

心理学导读系列

【美】Frederick T. L. Leong , James T. Austin 主编

周晓林 訾非 黄立 穆岩 等译

心理学 研究手册

The Psychology
Research Handbook

A Guide for Graduate Students
and Research Assistants

 中国轻工业出版社

与那些充满高深莫测的内容、令研究新手望而却步的研究手册不同，
本书将激起你对心理学研究的终生眷恋。

——摘自《心理学研究手册》前言

正如著者在其“导言”中所说的，导师的时间和精力与学生的培养要求之间总是存在矛盾。一本《心理学研究手册》篇幅不大，却简明而全面地向学生和研究助手提供了研究工作所必需的基本技能，提醒读者在实际工作过程中应注意的事项，介绍一个独立科研人员必须掌握的拓展知识。

——摘自《心理学研究手册》译者序

封面设计：黄金支点

ISBN 7-5019-5079-2



9 787501 950799 >



ISBN 7-5019-5079-2/B · 078

定价：40.00 元

心理学导读系列

The Psychology Research Handbook
A Guide for Graduate Students and Research Assistants

心理学研究手册

【美】Frederick T. L. Leong, James T. Austin 主编

周晓林 訾 非 等译
黄 立 穆 岩

 中国轻工业出版社



图书在版编目(CIP)数据

心理学研究手册 / (美) 梁 (Leong, F. T. L.), (美)
奥斯汀 (Austin, J. T.) 主编; 周晓林等译. —北京:
中国轻工业出版社, 2006.1
(心理学导读系列)
ISBN 7-5019-5079-2

I. 心... II. ①梁... ②奥... ③周... III. 心理学研究
方法—手册 IV. B841-62

中国版本图书馆 CIP 数据核字 (2005) 第 099234 号

版权声明

The Psychology Research Handbook
Copyright © 1996 by Sage Publications, Inc.
Published by arrangement with Sage Publications, Inc.

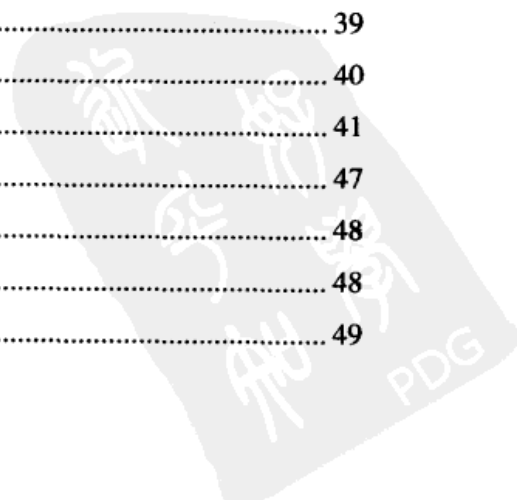
总策划: 石 铁
策划编辑: 张金颖
责任编辑: 朱 玲 张金颖 责任终审: 杜文勇
版式设计: 时春雨 责任校对: 万 众 责任监印: 刘志颖

出版发行: 中国轻工业出版社 (北京东长安街 6 号, 邮编: 100740)
印 刷: 北京天竺颖华印刷厂
经 销: 各地新华书店
版 次: 2006 年 1 月第 1 版第 1 次印刷
开 本: 787 × 1092 1/16 印张: 23.00
字 数: 400 千字
书 号: ISBN 7-5019-5079-2/B · 078 定价: 40.00 元
著作权合同登记 图字: 01-2005-0766
读者服务部邮购热线电话: 010-65241695 85111729 传真: 85111730
发行电话: 010-65128898 传真: 85113293
咨询电话: 010-65595090, 65262933
网 址: <http://www.chlip.com.cn>
E-mail: club@chlip.com.cn
如发现图书残缺请直接与我社读者服务部 (邮购) 联系调换
50908J6X101ZYW



目 录

第一部分 研究计划	1
第 1 章 寻找研究主题	3
搜索策略	4
细化和提炼主题	11
推荐阅读	14
参考文献	14
第 2 章 有效利用图书馆	17
参考或研究	17
定位：个人图书馆、当地学院图书馆或虚拟图书馆	17
各种资源	21
推荐阅读	27
参考文献	27
第 3 章 进行文献检索	29
进行文献回顾的步骤	29
组织文献	34
总结	37
推荐阅读	37
参考文献	38
第 4 章 阅读和评价研究文章	39
目的和听众	39
蜻蜓点水的阅读	40
评阅一篇研究文章	41
总结	47
推荐阅读	48
参考文献	48
附录4.1 评阅注释范例	49

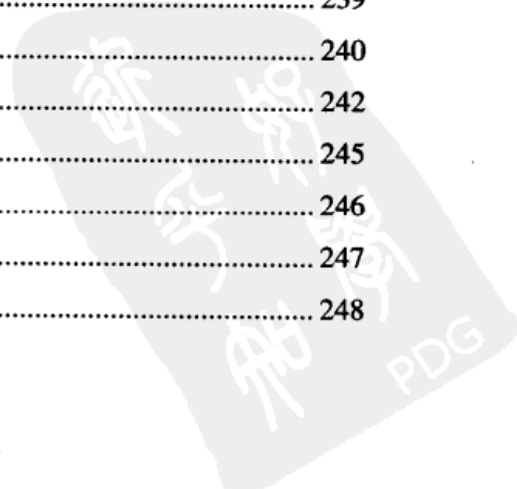


第二部分 设计、选择工具以及取样	53
第 5 章 设计一个实验研究	55
研究的目的	56
设计和推理过程	57
研究设计的类型	59
研究效度	62
总结	66
推荐阅读	66
参考文献	67
第 6 章 评估和选择研究工具	69
查找研究工具	69
回顾找到的工具	70
评估工具的心理测量学性质	71
效度	72
信度	74
什么是信度的满意水平	76
总结	76
推荐阅读	76
参考文献	77
附录6.1 工具和测验资源：书籍、期刊以及测验出版商	78
附录6.2 使用申请书样本	79
第 7 章 调查与问卷设计	81
调查、问卷设计与管理的基本知识	81
调查的设计问题	84
问卷调查的替代方法	89
总结	90
推荐阅读	90
参考文献	90
第 8 章 取样过程	93
取样和科学方法	93
非概率性取样技术	95
概率性取样	97
总结	102

推荐阅读	103
参考文献	103
第三部分 收集数据	105
第 9 章 申请用人类被试做研究	107
审查委员会的发展	107
IRB 审查的关键点	110
为审查委员会 (IRB) 的审查作准备	114
推荐阅读	115
参考文献	116
附录 9.1 俄亥俄州州立大学人类被试审查委员会行为和 社会科学研究概览表及调查者核对单	117
第 10 章 邮件调查	121
调查法的优缺点	121
抽样和返回率	123
附信	124
总体格式	125
问题设计	127
总结	129
推荐阅读	129
参考文献	129
附录 10.1 一份邮件调查清单	131
第 11 章 电话调查	133
电话调查的选用标准	134
产生样本组	139
人员	141
总结	144
推荐阅读	145
参考文献	145
第 12 章 从团体中收集数据	147
选择团体研究的环境和设计	148
数据收集程序	150
团体研究的数据分析	153
总结	154

推荐阅读	154
参考文献	154
第四部分 数据分析	157
第 13 章 数据预处理及初步分析	159
数据预处理	159
初步分析	164
总结：整理数据需要吃一定的苦	166
推荐阅读	167
参考文献	167
第 14 章 定性分析	169
定性研究范式	169
定性研究的策略	175
方法论上的共同点	175
定性数据分析的具体细节	176
总结	181
推荐阅读	181
参考文献	181
第 15 章 统计研究和探索的基本指南——统计分析和选择	183
收集数据前的分析处理	184
假设评估前的分析处理	184
检验假设	187
一般线性模型	190
回归	191
建模和回归	191
多变量分析	194
总结	194
推荐阅读	194
参考文献	195
第 16 章 基本统计分析	197
开始	197
描述性统计	198
推断性统计	199
总结	204

	推荐阅读	205
	参考文献	205
第 17 章	高级统计方法的使用	207
	效力	207
	变量之间的相关度	209
	组间差异的显著性	209
	结构	212
	总结	214
	推荐阅读	214
	参考文献	215
第 18 章	进行元分析	217
	元分析简史	218
	元分析的过程	220
	总结	224
	推荐阅读	225
	参考文献	225
第 19 章	测验的编制	227
	为什么要编制一个测验	227
	评估测验质量	228
	测验编制程序	229
	量表编制的其他方法	234
	总结	236
	推荐阅读	236
	参考文献	236
第 20 章	使用档案数据	239
	档案数据是什么	239
	什么时候应该考虑使用档案数据	240
	档案数据的获得	242
	研究过程的步骤	245
	总结	246
	推荐阅读	247
	参考文献	248



第五部分 研究报告的写作	249
第 21 章 APA 文体写作——为何与如何	251
研究报告的组成成分	252
文字处理与 APA 风格	254
总结	255
推荐阅读	255
参考文献	256
附录 21.1 APA 格式写作示例	257
第 22 章 打草稿	265
奇妙的思维	265
主人翁精神	266
下笔如涌泉	267
搁置	268
修改	269
听从建议	269
总结	271
推荐阅读	271
参考文献	272
第 23 章 修改研究报告	273
审稿者是什么人	273
收到评审意见	274
修改研究报告时通常要涉及的部分	276
推荐阅读	281
参考文献	281
第 24 章 应对杂志编辑和评审人	283
准备文稿	283
选择一个发表途径	283
编辑的一些典型反应	284
修改稿件	286
应对苛刻的反馈	286
与编辑建议的不同意见	287
推荐阅读	288
参考文献	288

第六部分 其他主题	289
第 25 章 协调研究团队——良好实验室的保持和发展	291
主动性	291
团队领导	292
沟通	293
团队的研究规范	294
研究助手的招募	294
训练和可靠性检查	295
实验室手册	295
实验室会议	296
数据检查	297
实验室压力	298
实验室的管理层次	299
维系一个良好的实验室	299
总结	300
推荐阅读	300
参考文献	300
附录25.1 实验室手册的样本内容	301
第 26 章 工作方式的多样性	305
不同类型的差异	305
心理差异的起源	306
理论和测量	307
了解自己的偏好	308
应用范围	309
成长和变化	316
提示	317
推荐阅读	318
注释	319
参考文献	319
第 27 章 申请研究基金	321
为什么要写研究基金申请	321
你寻找, 你就会有所发现	322
获得资助的小窍门	327

	推荐阅读	328
	参考文献	329
第 28 章	跨文化研究方法	331
	跨文化研究的特点	331
	文化对等	332
	跨文化研究的设计	333
	对文化的抽样	334
	被试的抽取	334
	数据收集	335
	跨文化研究的四种常见类型	336
	总结	336
	推荐阅读	337
	参考文献	337
第 29 章	在研究中应用理论——科学理论和科学研究的相互影响	339
	什么是科学理论	339
	研究想法的产生	342
	科学工作的循环	343
	对“在研究中运用理论”的进一步讨论	344
	总结	346
	推荐阅读	346
	参考文献	346



译者序

我正式指导研究生已经5年多了。作为导师，这5年应该说是一个痛苦的过程，一方面是因为自己缺乏学生指导经验，对许多问题不知如何处理；另一方面是因为学生的科研动力、基本素质参差不齐（即使是在北京大学这样的“名牌学府”）。虽然导师的职责主要是指导学生的研究工作，我却不得不花费额外的时间为这些学生“打工”，一而再、再而三地为他们补大学、甚至中小学的课程（不是夸张，现在有多少研究生能不经帮助而写出文法正确、前后连贯的中文句子，更不用说做到结构严谨、条理清晰）。其中最大的困扰就是，不管是心理学科班出身，还是其他专业半路出家，大部分研究生缺乏从事研究工作的基本训练，对科研结果和结论的产生过程知之甚少。他们常常不能适应从被动的、接受性的本科教育到主动的、创造性的研究生生活的转变，面对大量的文献、导师纲领性的要求以及需要自己动手，组织、实施研究方案的任务，常常表现出茫然不知所措。

问题不解决的话，这些困扰会随着我国心理学事业和研究生教育的发展而越来越严重。据不完全统计，2004年全国共招收心理学本科生6000多人、硕士研究生1500多人、博士研究生220多人。这么庞大的队伍却因为研究方法、基本技能的训练不足而发挥不了应有的作用；同时，也给大学老师、研究生导师增加了许多的压力（有个别地方，一名教授要同时指导50多名研究生）。怎样才能既高质量地、又相对便捷地培养学生呢？

《心理学研究手册》一书为我们提供了部分答案。正如著者在其“导言”中所说，导师的时间和精力与学生的培养要求之间总是存在矛盾。一本《心理学研究手册》篇幅不大，却简明而全面地向学生和研究助手提供了研究工作所必需的基本技能，提醒读者在实际工作过程中应注意的事项，介绍一个独立科研人员必须掌握的拓展知识（如何写作论文、如何申请资金、如何领导一个团队、如何在研究中应用理论）。

应该说，本书介绍的许多内容在心理学本科教育中都有所涉及，但这些内容往往分散在不同的课程中，一般是作为知识，而不是作为如何应用于实际研究的技能而传授。本科生学习这些内容时，大都采取就事论事、甚至是应付考试的态度。因此，他们一般不能把这些内容融会贯通，更不会把所学的知识应用到实际工作中去（事实上，大部分高校的本科教育并不注重对学生研究素质和实际动手能力的培养）。因此，许多学生在学了4年的心理学之后，并不一定真正知道心理学研究是怎么回事，做本科毕业设计和论文时或在实际工作中才发现自己一知半解，书到用时方恨少。事实上，甚至是一些高级

学者的研究、调查结果也因方法论上的缺陷而显得苍白无力。例如，盖洛普公司对2000多人的抽样调查就可以准确地预测美国总统大选的结果，而我们对诸如“中国人有几个性伙伴”、“抚养、教育一个孩子到底花多少钱”之类的问题的调查结果，即使有几万人的数据支持仍招来非议，原因之一可能就是此类调查没有运用分层抽样的基本原则。

对本书的阅读和参照可以在很大程度上弥补这些不足，因为它以实际的研究过程为主线，简明而全面地介绍了从发现研究主题、编制研究计划、进行实验设计，到收集数据、分析数据和写作论文这一系列操作中的关键性技能和注意要点。在阅读这本书后，不管是本科生还是研究生，不管是心理学专业还是其他相关专业（如社会学、教育学）的人员，均可以把握研究的脉搏，掌握实际动手的技巧；年轻教师阅读这本书，则可拓展独立工作的能力；研究生导师教会研究生学习、使用这本手册，则可让自己从繁琐的劳动中解脱（我就曾多次应邀给学生讲写作论文的基本注意事项），有更多的时间和精力去关心学生研究工作的创新性。事实上，正是基于这些原因，我才组织翻译这本书。我要求这本手册成为自己的研究生必读、必备的工具书、参考书。

本书的翻译工作在两年前就已开始了。10多名研究生分头工作，付出了相当的劳动。但不出所料，这些翻译初稿错误百出，多亏了瞿非（博士后）、黄立（直博生）和穆岩（本科生）参加，事情才有了转机。瞿非、黄立、穆岩花费了大量的精力重新翻译、校对，我的秘书潘钰在此过程中也做出了贡献。即便如此，我也不得不花费较大的力气翻译、审校书稿，付出的劳动远远超出了我的预期。当然，误译之处依然难免，责任在我。

总之，我希望本书中文版的出版有助于中国心理学以及其他相关学科的本科生、研究生科研素质和技能的培养，有助于导师、教授科研工作地开展。说得直白一点，如果这本书售出的数目每年超过2000册，那在中国做心理学导师的痛苦就可以减轻很多。

周晓林

北京大学脑科学与认知科学中心主任

北京大学心理学系教授

2005年5月



前 言

《心理学研究手册》应成为那些渴望掌握心理学研究方法和步骤的人员的标准参考书。本书的编者Frederick Leong博士和James Austin博士正是为学生们提供了这样一本综合性用书，用以指导针对人类被试进行的多种心理学研究。他们的辛勤劳动应受到赞赏。

全书条理分明，脉络清晰：从识别研究主题到申请研究基金，在心理学研究的世界里，读者将进行一次充满智慧和挑战的趣味之旅。本书共29章，其中几章由富有经验的研究生协助完成。所有的作者在他们各自的领域内有着大量的专业知识和研究经验；所有的章节在各种研究步骤中都含有详尽的“脚本（或模板）”。这种与真实世界的衔接增强了许多章节的清晰度和实用性。

此书的另一个优点是，章节编排从开始直至结束均以实验研究过程为蓝本。这种编排使得这本手册成为心理学本科生学习独立设计、研究的理想教材。不仅如此，本书因每一章独立成文，使得它也适用于那些不需系统学习、只从章节中摘选所需知识的读者。而且，所有的章节组成了6个模块——若需要获得在研究设计和工具选择方面的帮助，可参照手册的第二部分；若需要学习如何进行研究方面的协作，则可从第五部分获得有效的建议和指导。

与那些充满高深莫测的内容、令研究新手望而却步的研究手册不同，本书将激起你对心理学研究的终生眷恋。本书的特别价值还体现在那些一般教程遗漏的章节，包括跨文化研究、与期刊编辑和评审交流、进行元分析以及如何使用数据档案。

依我之见，这本书是向学生和研究助理介绍心理学研究的首选。

Anthony J. Marsella



导 言

由于许多心理学研究具有团队合作的性质，所以在安排时间培训本科生、研究生和研究助理时，许多学术带头人陷入两难境地：一方面，同这些研究助理一起工作很重要，因为他们是未来研究的后备力量，有价值的交流对教员和学生来说都有好处，同时也因为研究管理是教员的责任所在；另一方面，当同时存在很多其他重要的工作时，指导研究助理、进行基本的研究技能训练便不得不与之竞争时间和精力。而且，许多本科生并不会长久地与一名教员呆在一起，而是不断更换实验室，以获得更多样的经历。如果教员不完整地把基本的研究工作交给研究助理，而是事必躬亲，那时间安排上就会有困难；反过来，如果将工作交给缺乏必备知识或技能的研究助理，指导改正他可能要比自己做这些工作更加费时费力。同样的问题也存在于当心理学教员面对许多很聪明但没有研究经验、希望有机会独立学习或研究的本科生时。

对此两难困境的解决办法包括：对研究助理进行小组指导；让有经验的研究生指导刚刚起步的研究生和本科生；在选择研究助理时将专业课程和研究技能作为一项考核标准。另一种解决办法是，提供一本适合于初级水平而不是通常所见到的以专家水平为导向的手册。虽然研究手册通常会涉及某个领域的专业知识，但基本的研究技能也可以以手册的形式加以传授。Miller的“研究设计和社会化测量（1991）”就是这样一本手册，只不过更适用于高级水平。

因此，这本手册为上述两难境地提供了一种解决方案。不同领域的研究者负责本书不同的章节，将研究作为一项技能进行全方位的阐述。这种解决方案将研究作为一个模板，使导师要求他们的助理在进行研究之前购买这本手册。这本书也可以作为系统学习研究方法的必要或补充的教材。本书针对的读者是心理学研究生和本科生，但负责大型研究团队的教员可能也希望拥有这样一本参考书。最近一项调查显示，心理学统计、方法论和测量方面的研究生反映：“20年来统计和方法论的课程设置几乎没有进步（Aiken, West, Sechrest, & Reno, 1990, p.721）。”

与其他书籍不同，这本手册使用了一个更为全面的框架来介绍研究过程，从研究计划开始，依次介绍了设计/工具选择、收集数据、数据分析、研究写作以及研究小组如何工作等。其他同类手册仅仅集中于研究过程的一个或两个方面，而且通常只有少数几个作者。比如，有些书讨论科学写作（Pyrzack & Bruce, 1992; Sternberg, 1993）；美国心理学会（APA）出版指导用书，涉及图书馆使用（Reed & Baxter, 1992）和APA格式

(Gelfand & Walker, 1990); 其他一些书则介绍了实验设计和统计中的一些问题 (Bruning & Kintz, 1987; Sommer & Sommer, 1991; Yaremko, Harari, Harrison, & Lynn, 1986)。这本手册的主要贡献在于, 在基础水平上对研究进行了全面的、系统的介绍, 同时又体现研究观点上的百家争鸣。各个章节的作者均有着研究的实践经验。

在这里, 我们先预览一下手册的各个章节, 这些章节将研究过程看作是一个认知脚本而统一于一个框架之中。认知脚本就是知识结构, 有时被称为“线性序列”, 它将如何做事的程序按时间顺序组织起来 (Smith, 1994)。我们认为, 专家和新手的一个主要区别是, 专家具备进行研究的更全面、更专业的脚本。最近的研究表明, 专业水平不断提高的人 (如本科生、研究生新生、初级和高级教员) 其研究知识体系不断得到完善 (Hershey, Wilson, & Mitchell-Copeland, 1996)。《心理学研究手册: 研究生和研究助理指导用书》将使心理学本科生、研究生、研究助理获得详尽和富有实效的各种研究脚本或模板。总之, 这本手册可以帮助教员脱离在指导研究助理或学生进行研究时所面临的两难境地。

我们按研究计划、设计和工具选择、数据收集、数据分析、研究报告的写作和特别主题来组织章节。这6个部分中的每一个部分都包含多个章节, 分别由富有实践经验的研究者以基础水平为着眼点进行撰写。

章 节 概 略

我们可以把研究计划看作是一个从寻找研究主题到有效使用图书馆、查阅文献、不断改进和完善研究论文的系列过程。在第一部分“研究计划”中, Leong和Pfaltzgraff (第1章) 对寻找研究主题进行了阐述, 并提供了一个应用框架, 解决如何使用个人化的、交流性的、计算机方面的以及书面的信息来确定主题。Reed和Baxter (第2章) 介绍了如何利用学校图书馆资源。Stockdale和Kenny (第3章) 展示了在确定研究计划时查阅文献的复杂和巧妙之处。最后, 收集到的文献必须经过批判性的评价, 这正是Oleson和Arkin所负责章节 (第4章) 的主题。

第二部分“设计、工具选择以及取样”中, 涉及的内容是如何检测自己的假设。我们所选的实例是设计一个实验研究 (Wampold, 第5章)、评价和选择研究工具 (Ponterotto, 第6章)、设计问卷和调查表 (Goddard & Villanova, 第7章) 以及实施取样过程 (McCready, 第8章)。

第三部分集中讨论了数据收集, 这是在论证假设过程中的一个关键步骤。具体过程包括向人类被试委员会申请并取得同意 (Schmidt & Meara, 第9章)、进行邮寄方式的调查 (Vaux, 第10章)、通过电话进行调查 (Chen, 第11章) 和收集群体数据 (Zaccaro & Marks, 第12章)。

第四部分关注数据分析。统计方法为研究者提供了一种在进行探索和求证研究中检验数据的工具。具体步骤包括筛选数据和进行预分析（Dollinger & DiLalla, 第13章）、选择统计检验方式（Yaffee, 第15章）、用基本的分析回答研究问题（Dickter & Roznowski, 第16章）、使用更高级的分析回答研究问题（Levy & Steelman, 第17章）。其他的步骤包括利用已有的研究进行元分析（Cooper & Dorr, 第18章）、以单个研究而不是个人为分析单位进行检测（Betz, 第19章），以及使用档案数据（Zaitzow & Fields, 第20章）。我们同时也介绍了定性分析（Highlen & Finley, 第14章）。

对科学家来说，写作就是对外交流，它可以使研究结果得到公开验证，与他人分享成果。在这种背景下，写作将涉及研究的具体细节，为研究提供连贯的阐述。在第五部分“研究报告的写作”中，Austin和Calderon（第21章）讨论了APA格式的基本原理和机制（根据APA手册第4版，1994）；Peterson（第22章）谈到了如何打草稿；Negata和Trierwiler（第23章）则讨论了如何结合前辈和同辈的意见对原稿进行修改。最后，Osipow（第24章）对如何与编辑和评审人员打交道提出了建议。

第六部分“特别主题”，将一些对研究工作有重要意义，但以前同类书籍中不予重视的主题集中在一起。这些主题包括如何协调研究小组（Molfese等，第25章）和了解工作方式的多样性（Brenstein, 第26章）。其他3个重要的主题是申请研究基金（Borkowski, 第27章）、跨文化研究方法（Leung & Van de Vijver, 第28章）、如何把理论应用到具体研究（Gelso, 第29章）。

在这本书中，我们提供了解脱困境的办法。需要着重指出的是，掌握这个过程的知识需要个人广泛的实践。因此，读者必须具体实践书中的建议。我们也希望读者在阅读各个章节时，试着举出实例，并尝试各类脚本。举例来说，批判性地评价文献，这个脚本必须结合几篇文章进行练习才能掌握。正如Peterson（本书第22章）所建议的那样，不要害怕得到别人的反馈，只有这样才能帮助完善自我评价。统计学尤其需要反复练习，对同一数据不妨进行不同的分析。写作也是如此。总而言之，主动地利用好书中的材料，将会帮助你真正掌握研究步骤和过程。



第一部分

研 究 计 划



第 1 章

寻找研究主题

Frederick T. L. Leong

Rhonda E. Pfaltzgraff

在本书的导言中，编者用了认知脚本（cognitive scripts）这个概念作为开展研究的框架。他们认为，技能的获得依赖于对脚本建立的练习。在一个线性的研究项目中，找到研究主题是最基础的认知脚本。其他后续步骤的认知脚本包括选择研究设计、选择研究仪器、选用抽样程序、进行数据分析等。寻找研究主题的技能是通过不断摸索逐渐发展起来的，但是学习其他人是怎样做的也很重要。本章旨在通过列举一些心理学家寻找研究主题的方法，向那些缺少经验的研究者说明如何“寻找研究主题”。

一个人是否对这样或那样的主题感兴趣，是一系列不同因素共同作用的结果；对这些因素，职业心理学专家已经作了长期的研究（如Osipow & Fitzgerald, 1996）。从某种意义上讲，寻找研究主题有赖于个人兴趣和爱好，就像对某种音乐或食物的偏好也是个人兴趣一样。考虑到在形成“感兴趣的研究主题”的过程中存在个体差异，对这一过程给予指导的最好方法应当是“菜单式方法”。仍以食物做比喻，如果人们对某种食物感兴趣的程度上存在显著的个体差异，那么，一个用多种方式烹饪食物、食品菜单丰富多样的饭店将有可能满足更多人的需要。同样的道理，不同的个体在寻找各自感兴趣的研究主题时也会有多种不同的偏好。这种态度也顾及了个体学习和认知风格的不同（如，Pintrich & Johnson, 1990）。读者也可以阅读Brenstein（本书第26章）关于个体差异如何影响一个研究小组的工作的论述。因此，我们在本章提供了一系列不同的寻找研究主题的策略。不管对研究策略的偏好存在多大的个体差异，我们很希望读者使用多种策略，而不仅仅依赖于个人偏爱的某种策略。就像爱因斯坦曾经说过的：“机遇偏好有准备的头脑。”想寻找一个重要的、感兴趣的研究主题，最好的方法是考虑一个可选择的、广泛的、范围，至少在研究的初始阶段应该是这样。寻找研究主题的最初步骤是一个全方位的过程，需要有对待新想法、新问题的开放式眼光，而非仅仅对已经很好发展的研究主题进行再加工。黏贴便条的发明者Art Fry，当他把其下属发明的、没有黏性的胶水与总是从教堂赞美诗集中滑出的书签相联系时（他对黏结具有开放的思想），一个

杰出的思想就产生了。其余的，正如人们所说，就是历史了。

采用不同策略寻找研究主题的另一个原因是时间紧迫。时间紧迫是很多研究项目要面临的问题。从硕士论文到作为助教的最初6篇文章，在人们试图寻找一个激动人心而又切实可行的研究主题时，常常会体验到紧迫感。使用“以问题为中心”的方法，能够帮助我们确定与我们相关且让我们感兴趣的课题。例如，某人的朋友或亲戚不幸患了抑郁症，这可能引发他把抑郁症作为心理健康问题之一的研究兴趣。但是，选定“抑郁症”作为研究主题后，还可以选择很多不同的方向。使用多种策略的方法有助于选定研究的方向。本章的其他部分阐述了寻找特定研究主题的多种策略。

寻找研究主题可分为两个步骤：(a) 搜索策略；(b) 细化和提炼主题。一些作者将第一个步骤解释为“对你想要从事研究的领域的全面认识”（Cone & Foster, 1993, p.28）。按照我们目前的计划，第一个步骤是寻找一个研究主题的范围，搜索策略有4类：

(a) 个人策略；(b) 人际间策略；(c) 印刷品资源；(d) 计算机策略（参见图1.1）。我们将依次讨论这几类策略。

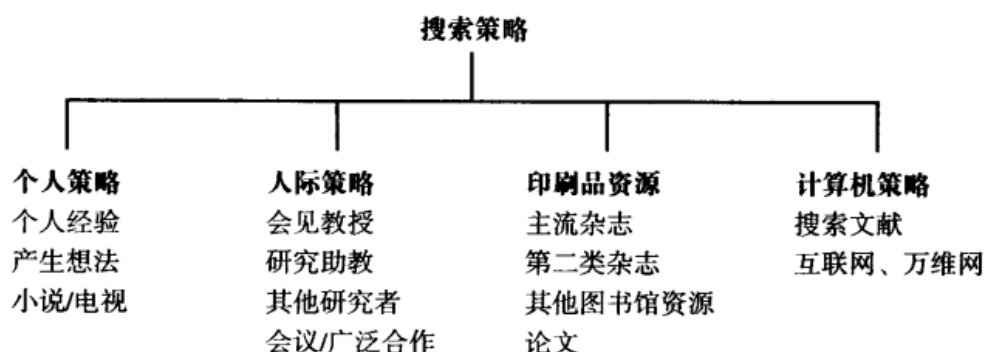


图1.1 寻找研究主题的搜索策略

搜索策略

个人策略

个人经验和观察一般是研究主题的来源，这对每个人来说都是易于得到的。很多心理学家通过个人经验来研究那些对他们有意义的人类行为模式。例如，一个非裔美国社会心理学家可能比一个欧裔美国社会心理学家更倾向于选择澄清种族偏见这个研究主题；一个在国外生活的临床心理学家可能更倾向于去研究精神分裂症表现的跨文化差异；一个作为独生子长大的心理学家可能对独生子女家庭和多子女家庭孩子的人格形成感兴趣。

个人观察同样可以产生新的见识。一个有机会长期观察人类行为的心理学家会很容易识别出潜在的研究主题。例如，一个咨询师注意到酗酒父母的孩子其注意广度较小，从而

可能引发探究其中关系的想法，以确定这是基于遗传还是基于环境的。虽然你不一定是个受过全面训练的心理学家，但是你也可以利用个人观察来产生研究思想。通过留意自己的观察并把它们写下来，你几乎可以不费力气地找到研究主题。研究思想可以来源于很多个人经验，例如，注意你的家庭成员如何抵抗疾病，或者观察你所知道的不同家庭中兄弟姐妹间的竞争，或者你会因为观看亨利·方达的电影《十二个愤怒的人》（*Twelve Angry Men*）而对团体动力学感兴趣，在这部电影中有单个人左右整个陪审团决策的情节。

当一项任务涉及面较宽泛并且其他策略都不起作用时，“头脑风暴法”有助于找到研究主题。这种产生思想的一般过程是：首先写下头脑中产生的所有主题，但并不立即判断每个想法的价值。在产生相当多的系列想法后，可以允许自己缩减其中的一部分，剪辑掉所有不切实际的或在自己手头的研究范围内不能实施的想法。如果你的任务是要写作有关儿童发展的书籍，那么写下所有你能想到的儿童赖以成长的过程，或随孩子的成熟而显现出来的各种行为。当你对自己的想法清单满意后，检查每个想法对这个任务的可行性、在该领域此研究的实用性和你坚持每个主题的动机。通过这样的方法筛选想法，你会选择出一个既可行又有趣的主题。

另一个产生思想创造性的策略是用“火星人的观点”作为认知策略。通过让自己想像一个刚刚降落到地球上的火星人会怎样看待某个特定问题或某个特定情境，你可能会发现其他人没有想到的有趣主题。例如，如果使用这种火星人的观点，我们就会很快发现，我们对汽车的迷恋已经远远超出了把它当作一种交通工具。并且，作为火星人，我们会注意到很多人宁愿通过对他人的身体或物质上的伤害来解决冲突，而不愿尝试去解决产生冲突的潜在问题。用这种探索方法，我们可以发现那些在心理学领域没有被全面探索的非常有趣的问题。

读小说、听广播、看电视同样可以产生研究主题。这些信息来源能揭示一些需要得到科学研究的帮助才能解决的问题。例如，当新闻讲述军人从武装冲突中返回适应平民生活所遇到的困难时，有人可能意识到，对创伤后应激障碍（Post-traumatic Stress Disorder, PTSD）的研究还没有对他们的适应问题提供充分、全面的解释。听过新闻广播后，一个对临床心理学感兴趣的学生可能想去更详细地研究这些人在适应平民生活过程中真实存在的和想像的障碍。个人搜索策略的主要目标是觉察和分析个人经验，从而识别出尚未解决的问题。

人际间策略

其他人也是研究主题很好的来源。其中重要的一个是大学教授。我们建议你将自己的教授作为想法来源之一，用一系列方法去汲取他们的知识和经验。如果一位教授目前的研究工作正处于你感兴趣的领域之中，那么与他会面将是非常有益的。很多大学教授熟知研究成果间的矛盾，以及目前可找到的文献证据上的差距。寻找可能的研究主题的另一条出路

是选修某个教授的课程。不管在上课时或是在课后，你都可以请求教授推荐你所在的研究领域中综述性的重要文献。你选择的课程范围应该从开设的普通公共课程到可以指导你个人兴趣和需要的专门课程。有一种指导型课程是为高年级本科生和研究生开设的专业阅读课程。阅读课程会提供在教授专长的特定主题领域接受个人督导的机会。利用教授这个资源的另一个方法是请求分享他或她的著作，这样你可以更多地了解教授的研究。

读过教授的著作或与他交谈后，可以讨论一下你是否可以成为他的助教。助教是寻找研究主题的绝好机会。作为助教，你可以熟知现有的研究主题，通过扩展教授的研究范围丰富著作的内容。通过当助教，你会很熟悉某一方面的文献，这可以帮助你找到契合兴趣并需要继续探索的研究领域。作助教的经历不仅为你提供熟悉教授研究项目的机会，而且如果你能够创造性地将教授的研究拓展到另一个水平或另一个领域，那么你就会有机会开展研究主题。例如，你帮助教授开展一个牵涉到滥用毒品的家庭动力学项目的研究，当对这个项目熟悉后，可以在教授发现的关系中探索其潜在的人种/种族和性别差异等。

除了大学教授，其他积极的研究者也可以为你提供新的研究思路。应用领域的研究者可以指出需要研究的重要问题。例如，他们可能会告诉你对目前公众看待艾滋病态度的看法。你可以继续这个想法，去研究态度的转变将会怎样减少对HIV病毒携带者的歧视行为。在商业环境中也有很多研究者。例如，为保险公司服务的研究者可能去研究人们对风险因素如何反应，或尝试鉴别能成为良好的保险销售员的人格特质。州政府和国家政府也雇用一些研究者，解决诸如公众健康或选举行为等问题。与校外的研究者电话联系或见面将为你提供额外的研究主题。

参加研讨会议和网络会议是另外两个寻找研究主题的有用策略。会议演讲以及张贴的海报可以让你大范围地了解目前正在被研究的想法。倾听或阅读其他人的想法可能让你找到研究主题。将别人的研究和你自己的想法融合，或将别人的研究扩展到新的领域是产生想法和寻找研究主题的有效策略之一。会议上的广泛合作或了解你所在学院的教授、学生同样是有益的搜索策略。注意别人的想法和目前感兴趣的领域有助于开阔你的眼界和思路。广泛合作可以帮助你熟悉那些和你有相似研究兴趣的人，这些人可能会建议一些潜在的研究领域。

印刷品资源

印刷品资源对于产生研究想法是有益的（参见 Reed & Baxter, 1992）。以前的科学杂志是研究主题最及时有效的和切题的来源之一。浏览最近的杂志可以发现目前备受关注的主题。略读一下相关杂志连续三四年来的内容列表和文章摘要。要确定去寻找最有意义的科学杂志，首先确定你最感兴趣的研究领域，然后找到在这个心理学领域备受青睐的杂志。例如，要寻找关于咨询心理学的想法，就可以选择《咨询心理学杂志》（*Journal of Counseling Psychology*）；关于工业和组织管理心理学，可以选择《应用心理学杂志》

(*Journal of Applied Psychology*)；关于社会心理学，可以浏览《人格和社会心理学杂志》(*Journal of Personality and Social Psychology*)。

那些不能确定对心理学哪个领域最感兴趣的人可以看看《心理学杂志》[*Journals in Psychology*, 美国心理学会 (APA), 1993]。这个出版物列出了在心理学各方面的主导杂志（在1993年第四版中超过了360个），并且提供了编纂原则的简短总结（例如，哪种类型的文章适合发表于这个杂志，这个杂志覆盖了哪些领域）。这个出版物还提供了一个非常有用的分类索引，帮助读者选择不同的主题领域（例如，咨询与临床、犯罪和司法、发展趋势）和不同的次级主题领域（例如，咨询与临床心理学类别下的健康心理学、发展心理学类别下的年龄老化问题）的杂志。然后，针对研究主题仔细地阅读。

另一个查看主要科学杂志的方法是浏览APA提供的专业摘要。为了服务于繁忙的研究者和临床医生，APA已经出版了一些针对特定主题的专业摘要期刊。例如，关于临床心理学，一类是关于应用心理学，另一类是关于心理分析。这些期刊是定期出版的，提供了一个快速便捷的方法来总览特定领域正在进行的研究。类似的资源，如《当前内容》(*Current Contents*)，虽在其他出版社出版的，但在大学图书馆里也可以找到。读者可以查阅Stockdale 和Kenny（本书第3章）来了解这些图书馆资源的更多详情。

主要的杂志资源有两种类型。我们已经指出，主要的科学杂志可以作为研究思想的来源，另一个来源是主要的临床杂志。很多杂志出版诸如个案讨论和陈述的临床信息，这对产生研究想法也是有益的。例如，这类杂志的某篇文章描述了几例男性病人患上不寻常的性恐惧 (Beit-Hallahmi, 1985)。具体来说，这些男性害怕女性的阴道是有锯齿的（即阴道齿状物）。这样，如果与女性发生性关系，他们就会要受到严重的伤害。按照心理分析的理论，阴道齿状物（或对阴道有锯齿的幻想）标志着男性的阉割焦虑 (Eidelson, 1968, p.465)。不管人们是否相信Freud关于阉割焦虑的概念，这些特例很好地说明，临床杂志如何包含很多有趣的人类问题，同时为产生研究主题提供了不可忽视的珍贵资源。

次要资源是那些为分发给其他读者而对首要信息材料资源所作的总结。其中之一就是教科书。在进行研究时，最好是参考首要资源，获取有关自己研究主题的研究成果以及这些研究是怎样进行的信息。诸如导论性课本和专业论文集都是产生研究想法的宝贵资源。即使你对自己的主题领域知之甚少，它们也是特别有用的。如果你选择的主题领域很广泛，例如“认知心理学的任何主题”，开始获得一个总体想法的绝佳途径就是那些涵盖了大多数认知心理学主题的认知心理学课本的目录。专业课本的总结部分对产生研究主题同样是有用的，因为它们让你对特定研究领域的问题和正面临的挑战有了一个快速的概观。

在确定一本教科书中你感兴趣的章节并去除你不感兴趣的章节之后，看一看在每个章节中列举的研究项目和主题领域，决定你愿意研究的主题。被书中章节引用的研究项目可

以帮助你很快地了解，确定深入研究这个主题领域是否仍然有趣。专业论文集，例如《心理学年鉴》（*Annual Review of Psychology*），同样有益，因为它们回顾一个主题领域并指出继续研究会有所收益的领域。这些文章也可能引用那些在该领域被认为是经典性的研究，并提供对你研究有用的背景知识。另一些能提供对于研究主题概观的专业论文集是图书馆中的各种手册和专业词典，这些是寻找特定领域的所有不同主题的绝好书籍[如Eidemberg的《心理分析专业词典》（*Encyclopedia of Psychoanalysis*）]。

除了导论性课本和专业论文集，另一种材料也属于次要信息资源。这包括学术报告、白皮书（国家健康研究院称之为“大多数人意见一致的论文”）和会议学报等。很多这些特殊的报告，可以在大学图书馆的“联邦政府收藏”部分找到，这都是研究想法的潜在来源。这些报告与首要的科学杂志一样，包含了在心理学各个领域的最新材料，通过给读者一个方便回顾的总结形式，从而为读者提供了更多的便利。例如，标题为《专业报告：精神分裂症1993》（*Special Report: Schizophrenia 1993*）的出版物（国家心理健康研究所，1993），总结了对精神分裂症所作的优秀研究。它长达90页，所总结的内容可能要花费一位精神分裂症研究者几年的时间去阅读和回顾。一个相似的信息来源和对行为医学的概观来自于报告《行为医学最新前沿》（*New Research Frontiers in Behavioral Medicine: Proceedings From the National Conference*，国家心理健康研究所，1994）。

除了上述的首要 and 次要研究资源，学位论文对寻找研究主题也是有用的。这种论文有些与众不同，因为它们不是出版物但被广泛提供。这些资源对于那些为做论文或其他类型的研究而处于确定研究主题阶段的学生很有帮助。通过检索其他学生写的论文，你可以对这些类型的研究主题的广度和深度有所了解。另外，很多论文包含了文献综述部分，常常包含着关于未来研究的想法，就像特定心理学领域的详细综述一样。这些文献综述经常包含很多在杂志文章的简短引言部分发现的各种研究的深层信息。很多大学图书馆提供在该大学完成的所有论文的收藏。你还可以搜索“国际学位论文摘要”，这是一个对在美国大学完成的多数论文的年度系列摘要。

计算机策略

搜索策略的最后项是应用计算机资源。不管我们多么有创造性，我们的顾问多么学识渊博，个人的和人际的搜索策略在记忆容量、综合性和速度方面是永远赶不上计算机的。我们通过输入特定的关键词，可以在计算机的信息库中搜索各种主题的相关信息（Stockdale & Kenny，本书的第3章）。在心理学领域里，最主要的计算机信息库是美国心理学会出版的PsycLIT信息库。

和其他的索引系统一样，PsycLIT开始是作为论文媒介将心理学中精选的杂志分离出来，编入索引，作为图书馆期刊出版发行。在1927年刚刚开始时叫作“心理学摘要”，随着个人计算机和光碟的使用逐渐增多，PsycLIT信息库后来发展成为一个更有效的计算机

版本的“心理学摘要”，这个PsycLIT在很多大学的图书馆都有提供，或是通过光碟形式，或是通过电话拨号链接服务（如DIALOG）。

当试图在心理学领域确定一个研究主题时，用PsycLIT作为开始是很有用处的。但是，你不要将研究局限在一个信息库，因为每一个信息库都局限在选定的一部分杂志内，不可能每一个信息库都能列出你最感兴趣的特定主题的文章索引。其他主流的信息库如教育杂志（ERIC）、医药和健康科学杂志（Medline）、社会学杂志（Sociological Abstract）等，对于心理学研究是有用的。例如，你有兴趣探索行为药理学的一个主题，同时搜索PsycLIT和Medline信息库，要比你只用一个信息库获得的这类研究信息更全面。

除了很多大学图书馆提供的主流信息库以外，还有一些你应该知道的其他的专业信息库。列举出全部的专业信息库对我们来说是不可能的，因为其中有很多没被广泛使用。但对你来说，很重要的是意识到，除主流信息库之外，还有专业信息库，可能与你的主题相关，你应该在大学图书馆里搜索确认。

在这里，我们将阐释一下，综合应用这章讨论的不同策略是值得的。假设你对变态心理学和创伤后应激紊乱（PTSD）特别感兴趣，你可应用PsycLIT和Medline作计算机文献搜索，找到国立创伤后应激紊乱研究中心的研究者所作的相当数量的文章，这个中心属于美国退役军人管理局。通过查阅这些文章，你发现Matthew博士经常写这类文章，他是总部设在佛蒙特州White River Junction VA 医疗中心的国立创伤后应激障碍研究中心主任。你决定要运用一下人际搜索策略，安排一次与Matthew博士的电话会面来讨论你的研究想法。在交流中，他提醒你，国立创伤后应激障碍研究中心存有世界范围内有关这方面文献的信息库，并且他为你前去搜索这个信息库作了安排。通过这个专业的信息库，你找到了与研究主题相关的文章，它们比在PsycLIT和Medline信息库找到的要多得多。其他的专业信息库的例子包括由Dartmouth医学院留存的酗酒及毒瘾数据、由佛罗里达州甘尼斯威尔城（Gainesville）的心理类型应用中心所保存的关于Myers-Briggs类型指标（它是基于荣格理论的人格测试工具）的数据。

另一个有用的信息库是社会科学引文索引（SSCI）。SSCI索引是对某个专业文章的引用，而不是索引专业文章。例如，你对Bandura关于自我效能的理论感兴趣，可以使用SSCI信息库找到并引用Bandura最初表述模型的《心理学评论》（*Psychology Review*）。另外，如果你对Bandura形成自己的模型所作的研究感兴趣，你也可以使用SSCI找到Bandura这篇发表在《心理学评论》上的文章所引用的所有文献。SSCI的缺点是它的计算机版本比较昂贵，并且不是所有大学的图书馆里都提供。SSCI期刊形式的文字版本在很多图书馆都有，但它不具备计算机信息库的优点。

不管用哪种信息库，开始的时候最好备一本词典或手册。这些词典或手册将告诉你特定的信息库中摘选了哪些杂志以及用于索引这些杂志的关键词。当搜索一个主题时，使用正确的关键词非常重要。因为使用不适合的关键词，将从搜索中得到比预期要差的结果。

例如,你对职业目标(career goals)这个主题感兴趣,并用这个词组搜索了PsycLIT(1990—1995),你可能只得到很少的文章,从而得出结论说,在过去的5年内这类主题的研究很少。另一方面,如果你查阅了《心理学索引词词典》(*Thesaurus of Psychological Index Terms*, APA, 1994),你会发现,在PsycLIT中适合的关键词是职业期望(occupational aspirations)。使用这个关键词,你将发现,在1990~1995年间,有207个这个主题的相关研究,而不是用职业目标这个关键词时查到的38个研究。这样,明智之举是把一些时间用在词典或特定信息库的使用手册上,计划你的搜索,并保证你使用正确的搜索词。

词典或手册另一个重要的功能是,一种产生新想法的刺激来源。在翻阅词典时,你可以浏览心理学的各种概念和构想,从而在这些不同观点的分类中发现潜在的研究问题。例如,在浏览词典的过程中发现“职业期望”是搜索PsycLIT信息库的正确关键词,你可能想到心理学中一个有趣的概念“心理学上的雌雄同体现象(psychological androgyny)”,并怀疑是否在心理学上的雌雄同体与职业期望间存在相关:男性化的妇女是否与女性化的妇女有不同的职业期望。带着这些不同的概念和关键词,我们就可以搜索探讨过这些问题的研究。

除了这些信息库,计算机还可以通过另一个途径对寻找研究主题起到有益作用——这就是互联网。可以找到一系列的互联网使用手册(如Wiggins, 1995)。互联网中很大一部分是新闻或电子版的讨论集锦。有很多这些集锦值得你深入思考并作为研究想法的来源。例如,如果你对亚裔美国人的民族认同感兴趣,你可以登录Soc.Culture网站,并发现一些亚裔美国青年和中青年面临的问题(如在民族间融合的优缺点,在多大程度上美国化而成为所谓的“香蕉”,即外表是黄的而内在是白的)。另外,如果你对John Bowlby关于依恋的理论感兴趣,你可以参加一个重点在研究依恋理论的临床案例和潜在研究问题的专门讨论组。再举一个由“管理学院”主持的HRNet(人力资源)的例子,参加这个新闻和讨论组的人经常为不同的问题而贴出一些信息,寻求帮助。如果一个人贴出信息,请教网上是否有人了解关于民族认同的特定量表,其他人可能贴出信息,寻求与特定主题相关的文章。

最近网络技术发展很快,这是一个多媒体地址的集合,通过超文本将很多简易的入口链接在一起。超文本基于特定的计算机语言(hypertext markup language, HTML),从而允许使用者仅通过点击特定的、已设计好的文本,便可从一个信息源直接进入另一个信息源。你可以使用因特网进入各个心理学系的主页,从而知道不同院系的成员在研究什么,或者你可以登录国家心理健康研究所的主页,去查找正在提供哪种类型的研究基金。很多网页可以下载阅读文章,或从你的个人计算机中打印出来。例如,康奈尔大学在因特网上主持的“联邦玻璃天花板效应委员会”,链接了很多关于女性和少数民族遭受职业歧视的文章。很多因特网的浏览器和互联网网页具有“搜索引擎”,允许你搜索与感兴趣的主体相关的各种信息。这就是我们为什么能在网上找到“联邦玻璃天花板

效应委员会”。学习使用因特网的有效方法就是浏览最近几期《互联网世界》（*Internet World*），很多公共图书馆和大学图书馆都有这个杂志。

细化和提炼主题

在寻找研究主题的第一个阶段，你主要关注了如何确定一个大概的主题领域，诸如隐性抑郁症（*masked depression*）、女性中的工作满意度等问题。一旦你发现了自己感兴趣的领域，下一步就是细化和提炼主题。像Cone和Foster建议的那样（1993年），细化研究主题的方法，就是从你的一般主题领域中产生一个研究问题。他们指出，研究问题有三个特点：第一，它应该是一个问题，你应该把这个研究问题措辞为一个问题的形式（如男性和女性对电影暴力的反映一样吗）；第二，研究问题应该为待研究的变量建立一个潜在的关系（如具有较高自尊水平的青少年是否比低自尊水平的青少年较少可能染上毒瘾）；第三，研究问题应该使用可被实验检验的专业术语（例如，使用自尊水平而不是诸如“来自良好家庭的青少年”这样的模糊的概念）。

当你能够综合应用上述搜索策略来帮助自己产生一个一般的研究主题时，你也可以用同样的搜索策略细化和提炼你的主题，使之成为一个研究问题。例如，你最初研究主题的方向是天才儿童，你可以去拜访教育领域的教授，他们能告诉你这一领域的很多研究（个人搜索策略）。活跃在某一特定领域的教授通常会给你的后续研究指引一个大有可为的方向。在个人搜索策略之后，你还可以使用PsycLIT（计算机策略）做一个计算机文献搜索，来查找过去是否有关于出生顺序与天才水平之间关系的研究。另外，你可能想去继续查找印刷品资源，并开始阅读关于特殊教育的各种读本，以便在教育天才儿童这个主题中发现主要的研究问题。当你熟悉了这些主要的理论和文章后，可以继续应用计算机搜索策略，查找与你的研究主题相关的各类文章。不管你是否混合或匹配这些不同的搜索策略，在寻找研究主题这一阶段的目的，是把你的一般主题细化到一个具体的研究问题，它能够被实验检验，在特定的时间框架内切实可行，并有可能对科学知识做出有意义的贡献。

一旦你的研究主题细化到了具体的研究问题，下一步就是从你的研究问题中形成一个或多个假设（Cone & Foster, 1993）。因为科学的目的是理解、解释、预测和控制现象，心理学研究的主要功能之一是预测人类的行为。假设是对特定变量间关系的一种预测（如假设低自尊水平的青少年更易于对各种毒品上瘾）。检验和形成关于人类行为的各种假设就是心理学进步的过程。一般来说，我们使用零假设和研究假设来检验我们的问题。零假设是必需的，因为我们可以直接检验它们。零假设认为变量间是没有关系的，而研究假设与之相反，认为变量间是有关系的。例如，一个零假设可能是这样的：

“青少年的自尊水平与滥用毒品和上瘾之间没有关系”，而研究假设将是：“青少年的

自尊水平与滥用毒品和上瘾之间存在负相关”。除了界定变量（如自尊水平和毒品上瘾），假设还要界定变量间关系的性质（如正相关、负相关、曲线相关等）。

你的研究问题和假设都来源于我们上述列举的4种搜索策略。例如，你可以从读书（印刷品资源）来寻找一个大概的研究主题，然后由你作助教的教授帮助你找到特定的研究问题（人际策略）。另外，你可能使用PsycLIT信息库（计算机策略）查阅文章从而形成具体的假设。

最后，我们将提供几种概念框架或策略，在你细化和提炼研究主题过程中可能有用。在心理学领域，理解人类行为的一个广泛的概念图式是ABC模型。当使用这个模型时，心理学家首先选择一个他们最感兴趣的特定行为作后续研究。这些有趣的行为可以从学龄儿童的尿床现象到高压情况下的军队人事部门的功能。ABC模型将一个目标行为分为3个部分：前因（*antecedents*）——行为（*behavior*）——后果（*consequences*）。按照这个模型，一个目标行为有其特定的前因，或者说导致行为的原因以及特定的后果。如果尿床是目标行为，那前因可能是家庭压力，比如父母正在商议离婚，其后果之一可能是父母将注意力从相互的冲突中转移到孩子的高压体验上。

有时，使用ABC模型，从你已选择的大概主题中确定研究问题是很有用的。举前面提出的例子，假设你对女性的工作满意度这个主题感兴趣，你可以使用ABC模型来考虑女性工作满意度的各种前因。你使用这个模型搜索更多文献，发现在现有的关于职业的文献中常用的变量是传统职业（过于强调女性的性别）或非传统职业（不够强调女性的性别）。你可以把一般主题细化到一个研究问题，如职业选择的传统性是女性职业满意度的一个显著原因。你可以假设，从事传统职业的女性比从事非传统职业的女性有较高的职业满意度。

另一个细化研究主题的有用概念图式或策略是寻找潜在的中间变量。按照Baron和Kenny（1986）的说法，中间变量是“一个质的（如性别、人种、阶级）或量的（如奖赏水平）变量，它可以影响自变量与因变量间关系的方向或强度”（p.1174）。在相关分析中，中间变量是影响其他两个变量之间零相关的第三个变量。在方差分析中，中间变量代表的是自变量和因变量间的交互作用（更多细节参见本书Dickter & Roznowski撰写的第16章和Levy U Steelman撰写的第17章）。

最近的民意测验表明，美国黑人与美国白人对O. J. Simpson案裁决的理解存在差异，这揭示了种族是理解国家公安和司法系统的重要中间变量。其他研究者也曾尝试将文化描述为心理学中的主要中间变量。回到女性工作满意度这个例子，应用中间变量的概念（如“不是所有从事传统职业的女性都满意她们的工作”），你就可以注意到性别角色倾向性是这个研究领域的重要变量。女性职业选择的传统性与职业满意度的关系是否被她们的性别角色倾向性所影响？有关性别角色的文献指出，高度女性化的女性更倾向于进入或喜欢传统的职业，而男性化的女性则不易于满足这样的职业定位。应用中间变量

的概念，你可以提出一个研究课题，来检验女性职业选择的传统性与工作满意度间的协调关系。

另一个细化研究主题和确定研究问题的有用的概念图式是寻找可能的跨专业研究领域（如社会心理学和工业/组织心理学）。不同专业研究的主题之间经常有自然的或逻辑的关系，只需要很少量的创造性和创新能力就可以把它们融合并发展成一个值得继续探索的研究问题。例如，假设你对工业组织心理学、尤其是绩效评价中的性别偏见作为玻璃天花板效应（Morrison, White & Van Velsor, 1987）的潜在因素感兴趣。简单地说，玻璃天花板效应是一种对公司中女性职业发展仅因性别而遭受阻碍的假设。应用这种跨专业研究策略，你可以回顾过去几年社会心理学杂志所刊登的文章内容。在回顾过程中，你会发现社会认知领域是社会心理学中一个新的、令人激动的研究领域。你特别感兴趣于“组外同质性效应”（Mullen & Hu, 1989; Park, Jude & Ryan, 1991）。这个很强的效应是被社会心理学家发现的，指人类察觉到更强的组外同质性（如察觉美国黑人有更好的音乐才能，或者大部分女性开车水平差），以及更强的组内异质性。有人认为，组外同质性效应是刻板过程背后的机制。你意识到“组外同质性效应”可能是造成职业刻板印象的一个因素；你决定去测试这个假设：这个效应在绩效评定中起作用，并能真实地对别人报告的玻璃天花板效应中的性别偏差做出说明。

我们给你推荐的最后一个细化和提炼研究主题的概念图式或策略，就是寻找其他心理学领域的新构想。当应用跨专业研究策略时，你会发现，即使对你来说是新的概念和构想，在其他领域可能已经存在较长时间了（如社会认知）。当你试图在心理学领域寻找新的构想时，你既是试图找到最新发展的概念，也是探寻相关的测量手段。例如，咨询和临床心理学对种族间或少数民族间的寻求帮助或服务使用的模式已经有很多研究。很多过去的研究已经检验了诸如心理健康服务的可行性和临床或咨询师的理论倾向等变量。在最近的一个研究中，Leong、Wagner和Kim（1995）决定，检验心理学中的一个新构想——基于文化的“丢面子”这个概念——来作为影响亚裔美国人对寻求团体咨询态度的一个潜在因素。确认心理学中新构想的有效途径是回顾最近几期杂志（印刷品资源），同时参加新想法、新方法云集的各类会议（人际策略）。

在我们看来，构成寻找研究主题的策略有两个必须的要素。这两个要素就是对主题和创造性的掌握。不是说一个人必须对某一主题有特定水平的掌握才能进行重要的和有意义的新研究，当我们投入时间和精力学习研究主题时，也可以获得对主题事件的掌握（确定与主题相关的信息来源并学习它们）。

在掌握主题的基础上，找到和组织一个重要研究的另一要素是创造性。对创造性是一种人格特质还是可以被习得的技能这一点上虽然存在争论，但是我们相信，创造性多半是遗传和教养的融合。一些书对提高现有的创造力水平提出了许多建议（如，Adams, 1986; Sternberg, 1988; Weisberg, 1993）。一些文章提供了类似的建议，其中Wicker

(1985)的一篇文章堪称典范。花费一些时间来了解创造性的性质和过程以及一个人如何提高自己的创造力水平是很有益的,尤其是对那些有兴趣从事研究的人,这样他产生重要的研究想法的能力也会随之提高。

作为结论,我们希望上述提供的搜索策略及细化和提炼研究主题的概念框架,能对你在研究中寻找有意义的、重要的研究主题有帮助。

推 荐 阅 读

对如何寻找研究主题感兴趣的读者,可以阅读Cone和Foster(1993)的《心理学及其相关领域论文全集》(*Dissertations and Theses From Start to Finish: Psychology and related Fields*)。其他一些有用的资料可以在Long、Convey、Chwalek(1985)所著的《如何完成行为科学和教育学论文》(*Completing Dissertations in the Behavioral Sciences and Education*)一书的第2章,Dizon、Bouma、Atkinson(1987)所著的《社会科学研究手册》(*Handbook of Social Science Research*)“选择一个问题”这一章中找到。最后,由Campbell、Daft、Hulin(1982)所著的《怎样进行研究:产生和揭示研究问题》(*What to Study: Generating and Developing Research Questions*),对于寻找研究主题也是非常有帮助的(尤其是后两章)。

参 考 文 献

- Adams, J. L. (1986). *The care and feeding of ideas: A guide to encouraging creativity*. Reading, MA: Addison-Wesley.
- American Psychological Association. (1994). *Thesaurus of psychological index terms* (7th ed.). Washington, DC: Author.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182.
- Beit-Hallahmi, B. (1985). Dangers of the vagina. *British Journal of Medical Psychology*, 58, 351-356.
- Campbell, J. P., Daft, R. L., & Hulin, C. L. (1982). *What to study: Generating and developing research questions*. Beverly Hills, CA: Sage.
- Cone, J. D., & Foster, S. L. (1993). *Dissertations and theses from start to finish: Psychology and related fields*. Washington, DC: American Psychological Association.
- Dixon, B. R., Bouma, G. D., & Atkinson, G. B. J. (Eds.). (1987). *Handbook of social science research*. New York: Oxford University Press.
- Eidelson, L. (1968). *Encyclopedia of psychoanalysis*. New York: Free Press.
- Leong, F. T. L., Wagner, N. S., & Kim, H. (1995). Group counseling expectations among Asian Americans: The role of culture-specific factors. *Journal of Counseling Psychology*, 42, 217-222.
- Leong, F. T. L., Wagner, N. S., & Tata, S. P. (1995). Racial and ethnic variations in help seeking attitudes. In J. Ponterotto, J. M. Casas, L. Suzuki, & C. Alexander (Eds.), *Handbook of multicultural counseling* (pp. 415-438). Thousand Oaks, CA: Sage.
- Long, T. J., Convey, J. J., & Chwalek, A. R. (1985). *Completing dissertations in the behavioral sciences and education*. San Francisco: Jossey-Bass.
- Morrison, A., White, R., & Van Velsor, E. (1987). *Breaking the glass ceiling: Can women reach the top in America's largest corporations?* Reading, MA: Addison-Wesley.

- Mullen, B., & Hu, L. (1989). Perceptions of in-group and out-group variability: A meta-analytic integration. *Basic and Applied Social Psychology*, 10, 233-252.
- National Institute of Mental Health. (1993). *Special report: Schizophrenia 1993* (NIH Publication No. 93-3499). Rockville, MD: Author.
- National Institute of Mental Health. (1994). *New research frontiers in behavioral medicine: Proceedings from the National Conference* (NIH Publication No. 94-3772). Rockville, MD: Author.
- Osipow, S. H., & Fitzgerald, L. F. (1996). *Theories of career development* (4th ed.). Boston: Allyn & Bacon.
- Park, B., Judd, C. M., & Ryan, C. S. (1991). Social categorization and the representation of variability information. *European Review of Social Psychology*, 2, 211-245.
- Pintrich, P. R., & Johnson, G. R. (1990). Assessing and improving students' learning strategies. In M. D. Svinicki (Ed.), *The changing face of college teaching* (pp. 83-92). San Francisco: Jossey-Bass.
- Reed, J. G., & Baxter, P. M. (1992). *Library use: A handbook for psychology* (2nd ed.) Washington, DC: American Psychological Association.
- Sternberg, R. J. (1988). *The nature of creativity: Contemporary psychological perspectives*. New York: Cambridge University Press.
- Weisberg, R. W. (1993). *Creativity: Beyond the myth of genius*. New York: W. H. Freeman.
- Wicker, A. W. (1985). Getting out of our conceptual ruts: Strategies for expanding our conceptual frameworks. *American Psychologist*, 40, 1094-1103.
- Wiggins, R. W. (1995). *The Internet for everyone: A guide for users and providers*. New York: McGraw-Hill.



第 2 章

有效利用图书馆

Jeffrey G.Reed

Pam M.Baxter

参考或研究

由于各种各样的原因，研究者会使用图书馆来满足自己很多需求。这种使用可以分为两类：参考或研究。你可以使用图书馆满足即时的参考需求，如寻找一个事实或定义、查找某个人或某个机构、搜索一个概念或想法的概要、证明一个统计值；你还可以使用图书馆开展研究活动，查找资料来支持或反驳某个假设。与寻找具体的事实或某类信息不同，查找信息，开展研究，需要花费更多的时间和使用多种策略，并且不是每一种策略都可以成功。图书馆是否提供你要寻找的研究信息或它是否出版过该类读物也是无法保证的。虽然图书馆可以满足参考和研究两类需要，但在本章中，我们集中介绍图书馆如何满足研究者的需求。

定位：个人图书馆、当地学院图书馆或虚拟图书馆

大部分研究者使用几种图书馆：一种是个人在办公室或家里设置的个人图书馆，另一种是为学术机构服务的图书馆。如果研究者所在的地方有好几个收藏有趣图书的图书馆，那么可以利用的就不只是一种图书馆了。研究者如今正越来越清楚地意识到“虚拟图书馆”的存在，这是一种电子图书馆，它不是那种占据一定的物理位置、由一些工作人员保管和编档归类的充满书籍和杂志的图书馆。使用各种不同类型的图书馆有不同的优点和缺点（见图2.1）。

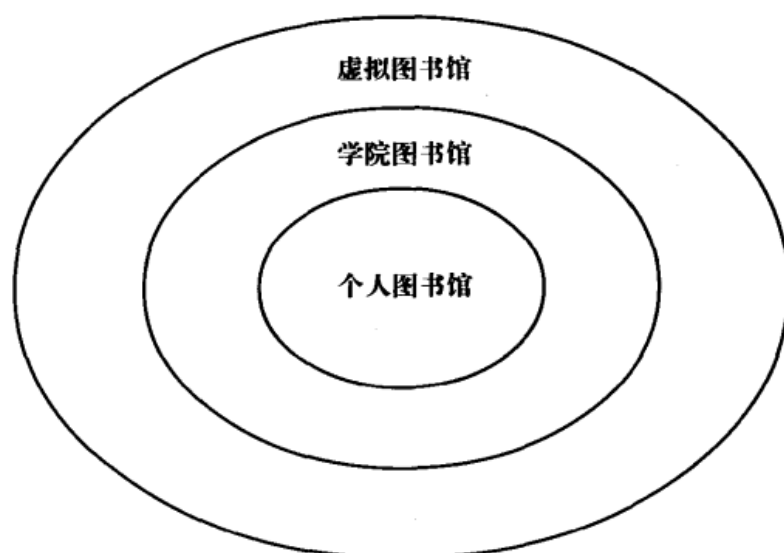


图2.1 按照与心理学研究者的远近展示几种图书馆类型

图书馆类型

个人图书馆

图书馆常与一定的物理位置有关，其中最频繁、最方便使用的是个人图书馆。个人图书馆通常收集你购买的专业书籍或订阅的杂志（下面将要谈到这些材料的类型和使用方法），还收集一些期刊文章的复印本、会议论文的复印本、从同事那里得到的属于你专业领域的其他材料。

使用个人图书馆具有明显的优点，因为在你的办公室就唾手可得，所以你不用担心图书馆的工作时间、你需要的材料是否正被其他人使用着等诸多不便，而且你知道自己的材料包含着哪些符合你和你同事研究兴趣的信息。

过度或单纯依靠个人图书馆满足参考和研究的需要也存在着不足。最明显的是花费，订阅杂志和最新的书刊手册是非常昂贵的。经费的问题可能影响手头信息的质量。为获得最新和最具权威的材料而花费的金钱和精力将影响你个人图书馆的规模和质量。

当地图书馆

相对于个人图书馆来说，你所在的学院图书馆（当地图书馆）能为你提供更多的便利。它的规模和收藏资料的范围比较大，并且易于查阅其他学科的知识，这对心理学中很多跨学科的研究是十分重要的。你可以依靠学院图书馆来购买过于昂贵的或与你的研究领域相差较远的资料；依靠图书馆工作人员所遵循的选择标准和程序，这些收藏图书可能比你个人图书馆收藏更新、更具权威性的资料；很多学院的图书馆欢迎和鼓励研究人员对购书提出一些建议。学院图书馆还提供一个重要的“增值”特色：那些提供服务的人员，他们可以扩展你对信息资源的使用。

就像个人图书馆一样,学院的图书馆也只能提供那些上架的图书。虽然大部分时间基本的参考书都被收藏在书架上,但是,还有一些因素使你无法获得资料。例如,你正需要的某一期杂志恰好在装订,某些书被其他人借走了,某些卷册已损坏,或者当你需要查资料时图书馆已关门。出于预算的考虑,会限制学院图书馆购书的频率和广度。很多学院图书馆无法获得特定类型的资料,如标准试卷或未出版的会议论文。最后,即使是很小型的学院图书馆,组织工作也并非易事:虽然你知道你拥有的每本书和杂志在个人图书馆的位置,但学院图书馆的藏书是按照预定的顺序编排组织的,对很多人来说这是武断的和难以理解的。

虚拟图书馆

虚拟图书馆最好定义为“存在的所有资料”,它不像个人图书馆或学院图书馆那样占据实在的物理位置。它包含很多其他图书馆的收藏,不管你是否知道它们的内容。幸运的是,很多图书馆通过拨号上网或局域网提供在线目录,你可以在它们的收藏中找到特定的书名。如果你需要某本书或某篇文章,无论这个材料是其他州还是其他国家所拥有的,这种馆际间的服务可以让你在当地的图书馆获得。

这种虚拟图书馆逐渐采取了电子格式,在网上提供电子目录,有相同兴趣的研究者开展电子讨论版或者为没有出版的信息提供研究资料(如电子版杂志或数据档案)。这种虚拟图书馆逐渐扩大的规模、范围和内容使得归类 and 建立目录几乎不可能。虽然保持时时更新更为不可能,但《互联网世界》(*The Whole Internet*, Krol, 1994)和《互联网地址簿》(*The Internet Directory*, Braun, 1994)列出了互联网的资源,更重要的是,它们提供了如何使你落于虚拟图书馆发展的建议。

因为你可以在家或办公室使用虚拟图书馆提供的各种便利,从某方面讲,它可以成为你个人图书馆的一部分。除了可能的链接收费或购买必要的硬件和软件,使用虚拟图书馆不需要直接的花费。由于电子资源的更新迅速、链接时间短、电子存储的花费少,使得即使只有少量的研究者使用也可以很经济。

像其他的计算机网络一样,使用过程中也会有一些不便,诸如下载时间、硬软件不兼容及上网接口等。在查找研究资料时,信息质量及版权是应多加注意的问题。许多电子杂志和及时更新的讨论版力争发布可信的和有权威性的内容,但是交流的速度和作者的多样性也产生了一些令人不快的结果,如对初步的研究结果就准予发表、未经证实的研究成果、不顾方法学和伦理学标准而给与所有的作者同样的评价以及欺骗性的和剽窃性的交流等。

使用各种图书馆的利弊权衡

如上所述,每种类型的图书馆——个人的、学院的、虚拟的——都有一定的优点和

缺点。另外，当研究者查找和使用信息资源时还要考虑其他的因素。

意外发现

当你翻阅一期杂志、浏览图书馆书架或查看即时提醒服务时，你可能出乎意料地遇到与你的研究相关的文章，或者与你设计研究时未曾考虑过的某些主题相关的文章。虽然这种方法不是完全可信赖的或可预测的，但在揭示许多你需要的意想不到的材料时，这些意外发现起着重要的作用。对材料的意外发现也指出了这种基于计算机的信息搜索系统的缺点，正像一个杂志文章的在线目录或计算机索引一样，特定材料的搜索结果必须使用你输入的关键词，这样就漏掉了那些与你的研究相关，但没有使用你所预期的关键词的文章。

无形的大学

很多研究者向同行传递信息的方法是出版文章、出书或撰写书籍的章节。另一方面，研究项目的很多部分并没有以实实在在的形式出现。这些包括负面的研究发现，使用某种类型的研究仪器或方法的困难（便于放弃一种而选择另一种更合适的），或在学术会议上有所陈述但未发表的论文中包含的材料。通常找到这类材料最好的也是唯一的办法即通过所谓的“无形的大学”，如休息时间咖啡壶旁边非正式的交流、非正式的学术聚会或学术例会上的论文陈述。在这类非正式接触中得到的材料可以成为你个人图书馆的一部分（比如，你索要一份论文陈述的复印件），或者通过虚拟图书馆这些材料可以为你所用（例如，通过电脑参加电子讨论的交流活动）。

非图书馆资源

有很多信息资源和版式在任何图书馆都很难找到，不管是个人的、地方的还是虚拟的。这些资料包括限量印刷的院系报告或研究报告、研究机构专有的或在网上不易得到的资料库，以及在标准的书籍目录工具中没有出现的回顾性的资料编纂物或专门收藏。另一个重要的但令人沮丧的资源是那些信息量不足而难以辨认的材料，比如作者在他或她的参考书目中给出的不完整或不准确的信息。

在这种情况下依靠一个专业的资料来源对于资料寻找过程是极为重要的，这样地方图书馆的“增值”特点便显示出来。对图书馆馆员提出的大量问题便包括这方面的内容。馆长工作的实质即辨别出资料是如何组织的，搞清楚现有的资料结构和跨度以及对你的研究需要和所需资料进行匹配。因为他们经常处理各种各样的文献，所以能对你的搜索策略提供广阔的视角。

各种资源

这部分将讨论在个人、学院和虚拟图书馆中可以找到的各类资源。对每种资源，我们将描述其目的和形式（即它是什么）、在研究中的用途以及存放的典型位置（个人的、学院的、虚拟的），并举出例子。如何寻找更广泛的资源列表以及使用特定工具的例子并未详尽列出，但可以参考Baxter（1993）、McInnis（1982）、Reed和Baxter（1992）、Sternberg（1993）的著述。本章我们将集中于工业/组织/工程心理学领域，并提供资料来源的例子。

教科书

目的和形式

教科书具有教育功能，通常为支持一定的教育计划而编撰的。指导性教材提供了某一领域如工业组织管理心理学的广泛概观（如Landy，1985；Muchinsky，1983）。它们总结了该领域的主要原则、理论、研究发现以及发展趋势等。中高级教材提供了诸如工作动机等子学科的更具体的研究方法（如Steers & Porter，1991）。

研究中的应用

每个好教材应该涉及与它所覆盖的主题相关的主要资源。因此一个教材可以提供对这一领域文献的入门路径。但应该认识到，教材是总结和报告其他人工作的次级资源。尽管它有利于提供对一个领域的总结和概观，但是作为次级资源也存在一定的风险，即可能曲解、过于简化或忽视了新的研究发现等。

位置

很多专业人员在个人图书馆中存有与每个感兴趣领域相关的导论性教材，以及与专业相关的高级教材。虽然图书馆不会去收藏大量的教材作为主要的资源，但是大部分学院图书馆还是存有有趣主题的一种或几种教材。你所在学院图书馆或虚拟图书馆的参考资料部门应该能够提供大部分教科书的信息。

主要专题著作

目的和形式

在每一领域都可以找到一些除教科书之外的长篇资料。经过一定时间后，有些专题著作成为每一个从事该领域的研究者都希望熟知的“经典”之作。有些经典之作成为本科生的教科书，并且对后来的心理学家产生了很大的影响（如Katz & Kahn，1966）。

研究用途

经典的专题著作在一定程度上对该领域提出了新的研究方法，它们成为首要的资源，并对后续的研究做出了铺垫。例如，Campbell、Dunnette、Lawler、Weick（1970）激励了许多关于经理人的研究，March和Simon（1958）的工作被认为是后来的有关组织理论、结构和冲突等方面工作的出发点。

位置

像高级教科书一样，研究者倾向于在他们的个人图书馆中保存几本主要的经典专题著作。因为一些重要的著作已经绝版并且很难得到，可能只能在学院图书馆中或偶尔在虚拟图书馆中可以找到这些著作。

手册

目的和形式

每隔一段时间，人们决定（并且得到出版商同意）筹备全面的参考书来总结一个领域的工作。一旦出版后，这种手册便提供了某一领域“科学发展动态”的全面评估，包括理论、研究成果、方法学和问题等。这种手册通常是多章节的，由多个作者编写，每位作者或专家撰写他或她擅长的章节主题。

研究用途

因为手册的章节提供了该主题的概观，所以它经常为这一领域的重要资料提供了广泛的参考书目。它可以为后续研究提供良好的开端。并且，手册会指出已经被研究的重要变量，某一领域手册的缺少表明这方面缺乏广泛的研究。

位置

《工业和组织心理学手册》（*Handbook of Industrial and Organizational Psychology*, Dunnette, 1976）在出版时被认为是对当时所做工作的重要总结。很多这一领域的专业人士索要复印本，存放在个人图书馆中。很多本科生用它作为工作和继续学习的重要资源。后来的第二版四卷对第一版做了更新和扩展，使得它在个人和专业图书馆中更具有价值（Dunnette & Hough, 1990—1994）。因为我们都面临财政上的限制，相关的或专业的手册 [如Bass 和Stogdill的《关于领导才能的手册》（*Handbook of Leadership*, Bass, 1990）] 通常在公共图书馆可以找到。虚拟图书馆也提供很多关于手册的资料或使用方法。

年鉴

目的和形式

综述展示的是关于某一领域主要方面的情况，是对最新进展有选择和有价值的回顾。例如，《心理学年鉴》（*Annual Review of Psychology*）从1950年开始每年出版一次，按照计划对这一领域的主题做定期的回顾。

研究用途

每一卷《心理学年鉴》包括20章，由该领域的专家撰写，每一个主题综述包括广泛的参考书目，为很多研究项目提供良好的开端。综述的一个优点是这种概要每年都出版并且比手册更新得更频繁。一些心理学领域有更为专业化的回顾与评论。

位置

一些学术专家会收到每一卷此类年度丛书来保证他们了解其他人的工作（如O'Reilly, 1991）。很多公共图书馆存有年鉴的所有卷册。但是，你可能必须依赖虚拟图书馆来查找更多的专业文章综述。

专家地址簿

目的和形式

这类资源提供专业人士的简短传记资料，包括目前的地址、工作地点及教育背景等。地址簿有很多种，大部分协会存有成员的地址簿（如美国心理学会成员地址簿）。

研究用途

通过提供地址信息，地址簿可以促进研究者之间的交流，并且通过罗列研究者的资历来表明他们的背景。

位置

在大部分协会中，成员资格的好处就是可以获得协会成员的地址簿。很多专业的心理学家在他们的个人图书馆中保存着所在协会成员的地址簿。公共图书馆可能存有国家主要组织部门的地址簿（如美国心理学会成员地址簿），但是没有所有协会的地址簿。很多公共图书馆倾向于集中成绩卓越的名人的地址簿，但不按学科分类，如美国名人录等。虚拟图书馆扩展了文字版本的地址簿的范围和广度。很多学院制作校园地址簿、电话簿以及便于在网上远程搜索的电子邮箱地址。随着因特网应用的不断延伸，现在一部分学术机构提供他们的教职员和研究者的详细资料，如他们的当前出版物的列表和目前

研究兴趣的陈述。

即时提醒服务

目的和形式

即时提醒服务提供某一特定领域或一系列领域最近出版文章的资料。这是至关重要的，因为很少有人有足够的资金去订阅所有可能相关的杂志，并且大部分人没有足够的时间去浏览该研究领域的所有相关出版物。

研究用途

即时提醒服务有助于研究者不落伍于该领域的出版物，例如：《当前内容：社会和行为科学》（*Current Contents: Social and Behavioral Sciences*）每周出版一次，并且再现了一年内1300多种杂志的内容列表。其他的服务项目集中于专门的学科领域。例如，APA 21分部的会员能收到PsycSCAN的季刊，即《应用实验和工程心理学》（*PsycSCAN: Applied Experimental and Engineering Psychology*）。这种服务提供选自100多种杂志文章的摘要，主要是关于计算机应用、人类动力学、人类工程学、安全学和工作条件等领域的文章。其他的PsycSCAN出版物覆盖了诸如心理分析、发展心理学、临床心理学等领域的内容。

位置

研究者可以在他或她的个人图书馆中得到这种服务，但通常更多的是在学院图书馆中。一些学院图书馆提供即时提醒服务，在收集了参与成员的兴趣后，他们将最近出版的杂志内容列表以硬盘拷贝或电子表格的形式复制出来。

杂志

目的和形式

心理学中的很多研究是以杂志文章的形式出版。心理学家感兴趣的有上百种杂志，即使在工业/组织/工程心理学领域中，按研究者的兴趣不同也有几十种可能的相关杂志。

研究用途

有能力的研究者在确定研究项目范围之后和开始收集数据之前，会回顾在他或她感兴趣的领域中以往的出版物。这样做是为了从过去的研究结果中学习，以确保研究变量的切题，将此项研究与目前的研究趋势结合起来，并确保这个项目过去未曾研究或报告过（除非其目的是重复某项特定的研究）。

位置

很多心理学家在个人图书馆中订阅几种最感兴趣的主要杂志[如《应用心理学杂志》(*Journal of Applied Psychology*)]。公共图书馆收藏更大范围的相关杂志[如《组织行为与人类决策过程》(*Organizational behavior and Human Decision Processes*)]。但是,公共图书馆也与个人图书馆一样,不能保存所有的相关杂志,尤其是那些过于专门的领域[如《团体动力学》(*Organizational Dynamics*)]。另外,从1994年起,一些杂志开始在互联网上提供电子版本。

摘要与索引

目的和形式

大部分索引和摘要提供杂志中搜索某杂志文章的作者、标题和主题。一些更专业的索引会出版会议论文和学报、书籍章节、本科生和研究生的论文、联邦和州的出版物以及技术报告等。它们能够被打印出来或者用计算机搜索。其中有一些只有一种格式,其他的则提供了好几种出版格式。

研究用途

其中一种用途是即时提醒。虽然它们和正式的“即时提醒”工具[如《当前内容》(*Current contents*)]相比更新速度慢了一点,但也包含了几千种杂志标题,并且提供了多种语言的文章摘要。另一个用途则是搜索过去的杂志文献,如印刷版的《心理学摘要》(*Psychological Abstracts*),甚至可以回溯到19世纪。

位置

直到1980年,大部分心理学家依靠印刷版的《心理学摘要》来进行文献搜索,或依靠图书馆管理员的服务来搜索PsycINFO信息库。渐渐地,获得这类资料的电子版格式更为容易了,研究者可以在办公室或在家里用计算机来完成搜索。一些政府出版的资料库,如ERIC,在公共网址已安装好并且可以在互联网上搜索。

论文

目的和形式

绝大多数的博士课程和很多研究方向的硕士课程都要求完成一篇论文。论文要求具有创新性的研究,即能够证明一个人在某领域的文献知识,可以使用合适的研究方法和工具组织研究,并且可以与别人就研究的内容和对该领域的贡献进行交流和辩论。

研究用途

因为论文包含了相关文献的广泛回顾，这对于确认过去的研究很有帮助。由于大部分研究既是提出、也是回答问题，所以一篇论文可对将来的研究提出建议。回顾过去的论文可以为完成自己的项目提供指导。

位置

学院通常要求学生提供论文的复印件，留存在大学图书馆中。对于博士论文，大部分还要求复印件要收录到UMI。UMI将收到的每一篇论文制作成缩影胶片，在《国际论文摘要》（*Dissertation Abstracts International*，或电子版的《在线论文摘要》）上宣布，并向个人及研究机构出售论文复印本。有时论文复印件可通过馆际借阅获得。

总之，上述资源如教科书、手册、综述有助于你确定和细化研究主题（参见本书第1章的详细内容），摘要和索引可以使你全面地回顾过去的文献研究（参见第3章的详细内容）。第4章将论述如何评价论文、文章、报告和你在研究中得到的其他材料的质量。

未来的图书馆

未来的图书馆是什么样的？学术研究和交流正在发生显著的变化，决定未来人们如何获得和使用信息。

- ◆ 印刷品资料的花费将继续增加，同时，提供的资料容量也继续扩大。如果按照目前的趋势，可能会导致个人和公共图书馆财政压力的增加，并且会迫使人们更多地依赖非本地的信息资源。
- ◆ 越来越频繁地出现电子出版物以及出版过程中电子媒介的应用。越来越多的资料以电子文件格式提供——刻录在光碟上的百科全书、年鉴、地图等。一些杂志出版商除了印刷品外，还使用电子杂志——甚至要用电子杂志代替印刷品。
- ◆ 源于研究者和研究机构的数据库可能越来越多地以电子格式提供。
- ◆ 使用互联网起源于1990年。像Mosaic和Netscape公司提供的图形浏览器，允许世界范围的快速电子传输，并且在提供可供研究者共享的各种类型的资料方面更加有变通性。
- ◆ 为保存起见而扫描过去的书面印刷品以及对复印件的要求，可能意味着要不断地把过去的文献变成电子格式。

在线全文资料的提供将意味着可利用到更多的信息。电子格式允许资料以更快的速度跨距离传输。但是，用户在使用虚拟图书馆时也将遇到很多障碍，例如，巨大的资料传输量使得有效的交流变得困难，并且从远距离站点获得资料将逐渐需要更多的费用。

推 荐 阅 读

《使用图书馆心理学手册》(*Library Use: A Handbook for Psychology*, Reed & Baxter, 1992)对如何有效搜索图书馆提供了很多附加信息。本手册包括读者使用图书馆搜索过程中需要的关键资料和指导的实例,目的是教会心理学系的本科生使用图书馆的技能。《心理学:参考和信息资源指导》(*Psychology: A guide to Reference and Information Sources*, Baxter, 1993)提供了几百种资料的参考书目,并附加了详细的注解。

另外,参考Krol(1994)和Braun(1994)、King和Kovacs(1994)的列表,它提供了上千种学术电子杂志和讨论版目录。

参 考 文 献

- Bass, B. M. (1990). *Bass and Stogdill's handbook of leadership* (3rd ed.). New York: Free Press.
- Baxter, P. M. (1993). *Psychology: A guide to reference and information sources*. Englewood, CO: Libraries Unlimited.
- Braun, E. (1994). *The Internet directory*. New York: Fawcett Columbine.
- Campbell, J. P., Dunnette, M. D., Lawler, E. E., & Weick, K. E. (1970). *Managerial behavior, performance, and effectiveness*. New York: McGraw-Hill.
- Dunnette, M. D. (Ed.). (1976). *Handbook of industrial and organizational psychology*. Chicago: Rand McNally.
- Dunnette, M. D., & Hough, L. M. (Eds.). (1990-1994). *Handbook of industrial and organizational psychology* (2nd ed., 4 vols.). Palo Alto, CA: Consulting Psychologists Press.
- Katz, D., & Kahn, R. L. (1966). *Social psychology of organizations*. New York: John Wiley.
- King, L. A., & Kovacs, D. (1994). *Directory of electronic journals, newsletters and academic discussion lists* (4th ed.). Washington, DC: Association of Research Libraries. [<http://www.mid.net:80/KOVACS>]
- Krol, E. (1994). *The whole Internet: User's guide and catalog* (2nd ed.). Sebastopol, CA: O'Reilly & Associates.
- Landy, F. J. (1985). *Psychology of work behavior*. Homewood, IL: Dorsey.
- March, J. G., & Simon, H. A. (1958). *Organizations*. New York: John Wiley.
- McInnis, R. G. (1982). *Research guide for psychology*. Westport, CT: Greenwood.
- Muchinsky, P. M. (1983). *Psychology applied to work*. Homewood, IL: Dorsey.
- O'Reilly, C. A. (1991). Organizational behavior: Where we've been, where we're going. *Annual Review of Psychology*, 42, 427-458.
- Reed, J. G., & Baxter, P. M. (1992). *Library use: A handbook for psychology* (2nd ed.). Washington, DC: American Psychological Association.
- Steers, R. M., & Porter, L. W. (1991). *Motivation and work behavior* (5th ed.). New York: McGraw-Hill.
- Sternberg, R. J. (1993). *The psychologist's companion: A guide to scientific writing for students and researchers* (3rd ed.). New York: Cambridge University Press.



第 3 章

进行文献检索

MARGARET S. STOCKDALE

THERESA KENNY

在一项研究中最缺乏乐趣和令人烦躁的工作恐怕就是检索文献和写文献综述了。如果你聪明绝顶、对事情满怀好奇心、是个求知欲强的读者和天才的作者，那么你已经比别人超前一两步了。本章针对那些仍然认为文献阅读是一项令人烦恼的工作的人写的。我们希望揭示文献搜索的神秘，并且通过打散和再组织，分为几个可操作的步骤，来证实文献搜索不仅是可以做到的，而且是十分有价值的。

本章提供进行文献综述的逐步指导，要牢记的是做这件事没有捷径可走。我们也提供书目式、百科全书性的、索引资料的清单以帮助你开始工作，同时，希望在如何使用这些资料方面对你做些指导。但是，我们还要再一次提醒你：图书馆管理员是你的朋友，他们的工作就是帮助你搞清楚如何找到和怎样使用各种参考资料。不要害怕提问！

最后，本章将对你如何处理已找到的所有文献提一些建议。也就是说，我们会告诉你怎样组织文献搜索工作。

本章不包括你需要知道的关于查找文献和写文献综述的所有知识。为了产生一个研究想法，我们建议你阅读本书由Leong和Pfaltzgraff撰写的第1章。为了不至于在图书馆的书架中迷失方向，我们建议你阅读由Reed和Baxter撰写的第2章。由Austin和Calderon撰写的第21章会帮助你了解APA的工作方式。你还应该阅读Oleson和Arkin撰写的第4章，它介绍了如何查阅和评价研究文献。要寻找进行元分析方面（一种特定类型的文献综述）的技术帮助，你还应该看一看本书Cooper和Dorr撰写的第18章。

进行文献回顾的步骤

初步决策

在你冲向图书馆（或登录互联网）之前，你需要确定三件事情：（a）回顾这种文献在你的研究中起什么作用；（b）你准备综述的范围；（c）你确切要找的东西是什么。

问题(a)和(b)比较容易回答,问题(c)可能要求你在确定研究主题之前先做一个初步的文献综述。

文献回顾的类型

几乎所有的学术报告都会包括文献回顾,但是不同类型的文献回顾服务于不同的目的。通常情况下文献综述作为研究项目的引言,需要你回顾过去的文献以强调你所作研究的范围。例如,你要研究一个关于目标设定中承诺的作用,那么,你就应该回顾关于目标设定和承诺过程的文献以及关于工作动机的基本研究和理论,因为这是目标设定应用的范围。文献回顾的另一种类型是研究本身作为综述。这样你需要尽可能多地查找有关主题的文献。如果没有找到全部的文献,你就应该尝试总结出其中的“真相”,并且就现有的文献做出评价。这种类型的文献可能发表在《心理学通报》(*Psychological Bulletin*)、《心理学年鉴》(*Annual Review of Psychology*),或出版为书籍及书籍的章节。虽然研究生的论文通常侧重于经验性研究,但年鉴文献综述部分应该是全面的、可独立成文的。所以要先问你自己,“我做的是为支持一个研究项目而作的特定的、高度集中的文献综述,还是要单独出版的全面性的文献综述?”

文献综述的范围

对于上述问题的回答会帮助你解决下一个问题:“我的文献综述需要多全面?”如果你要写一个单独出版的文献综述,你回顾的范围应该是有限的:查找关于你的主题的所有文献。这样的要求对于一个学期论文可能或者经常是很轻松的,但是如果你在一个团队中工作,这个团队想单独出版一篇文献综述,那么你还需要做一份穷尽的文献搜索。

有限的文献搜索显然是不能穷尽的,这样,确定你的搜索范围是很重要的。你的搜索以哪种方式来限定?你可能只感兴趣于特定时间段内出版的文献(如过去的10年中),或特定类型的文章(如在同行评论的心理学杂志上发表的文章、政府文件等),或特定的群体(如亚裔美国女性)。你应该确定搜索的范围,以保证不会浪费时间和资源去收集那些无用的材料。

主题和搜索特征

你的主题越具体、范围越狭窄,能够找到的文献就越少,这各有利弊。如果主题过于窄化,你可能找不到任何相关的材料;如果范围过于宽泛,你将被大量的材料所淹没。一个定义较窄的主题允许你深入地研究及书写文章,这经常是被科学杂志所推崇的。另外,如果你将主题定义得严谨而具体,你就能够做一个彻底的搜索而不会面临材料过多的困难。在你使用图书馆之前,要把一个主题定义得严谨和具体常常是有困难的。我们中的大部分人在集中于具体的问题前需要阅读广泛的主题。

开始搜索：从哪里开始？

进行全面的文献搜索包括几个步骤，并且需要查找不同类型的资源。为了对该主题有一个全面的了解，专题著作（例如书）和杂志文章都应该参考。因为心理学的主题经常来源于各种各样的观点——社会的、教育的、医药的以及组织的等等——要参考各种类型的资源以确保正确表述你的主题。

一般资料

正像第2章建议的那样，全面理解一个主题的有效方法是查阅一些讨论该主题的教科书和手册。这些讨论包括书目列表（通常在参考文献部分提供），这对增加你的材料是有益的。虽然大部分学院和大学图书馆不会收藏过多的教科书，但是一般来说至少会存有一种与学院课程有关的基本教科书。

为了查找有关自己研究主题的深层次的材料，你可能想去翻阅几本课本长度的专题文献。浏览一些合适的领域可以指导你找到论述某一专题的书籍。但是，要注意的是，很多心理学主题可以从不同的方面获得，对有关你的主题进行专题书籍搜索也很重要。例如，教育心理学方面的书专门介绍教育领域的专题；工业和组织管理心理学方面的书则侧重经济和管理的内容。

搜索专题著作分为两步。第一步是为搜索确定合适的关键词或专题题目。因为图书馆编辑好的书籍目录是面向所有人的，描述一个主题的关键词经常随时间而有所变化。你有时会发现，你所熟悉的或者看起来最合适的关键词并不是图书馆用来编辑该专题书籍的关键词。《图书馆专题题目》（*Library of Congress Subject Headings, LCSH*）对你选择最合适的搜索关键词提供指导，它是图书馆用来录入书籍时使用的按字母顺序排列的多卷专题列表。如果你最初输入的搜索关键词不是官方使用的专题题目，你会被指向使用正确的关键词。LCSH还可以为你提供宽泛关键词（BT）和相关关键词（RT），这对于你扩展搜索范围是有用的，还可以提供限制关键词（NT），这会有助于你更具体的搜索。

一旦你确定了可以使用的最好主题，第二步就是参考图书馆给出的书籍目录。这可能是一个卡片目录，更多和更经常的是在线（计算机化的）目录。任何一个单独的图书馆都不会涵括在LCSH上找到的所有专题，所以你需要搜索宽泛的和相关的关键词。还有，图书管理员还会帮助你进行地区的和全国的搜索，并帮助你通过馆际借阅获得某本书等。

具体的和最新的资料：杂志

虽然书籍会给你提供关于你的主题的深厚背景资料，但是这些资料常常不是最新的，

因为书籍的生产和印刷要花费一段时间。为了查找当前的相关研究，翻阅该主题的杂志文章是很重要的。查找杂志文章的工具是索引和摘要，这两种都给你提供一定时间段内某一主题的杂志文章列表，并附加一些必要的资料（例如引文），来让你识别每一篇文章。摘要还提供文章的简短总结。就像使用LCSH一样，没有一个图书馆拥有将每本杂志都覆盖的索引和摘要；这些工具告诉你哪些可以用来进行更广泛的搜索。还要强调一下，如果你找到感兴趣的、想要阅读的文章，但是图书馆又没有，你可以通过馆际借阅系统获得该文章的复印本。

像书籍目录一样，索引和摘要可能不是印刷品就是诸如光碟数据库的电子格式。电子索引的主要优点除了速度快，就是它可以对你的关键词进行跨越几年的搜索，而每一卷印刷版的索引只局限在一年内或更短的时间。虽然自动索引可以节约时间，但你要记住的是，这类工具没有印刷版本回溯的时间远，所以最好两种都使用。

除了使用索引和摘要来查找杂志文章，另一个有用的搜索技巧是浏览那些会出版相关该主题的几卷杂志的文章内容列表。并且，图书馆可能会订阅《社会和行为科学的当前内容》（*Current Contents in Social and Behavioral Sciences*），它会列出有关心理学的杂志即将出版的内容列表。你还可以向作者提前索要文章的复印件。

使用索引

在指出心理学领域中一些最有助于查找杂志文章的资料之前，我们先列出几种可用的基本搜索策略。当你查找某一专题的书籍时，使用正确的关键词至关重要。但查书没有像LCSH那样的资源能对所有的索引和摘要提供正确的专题题目。你可以先使用搜索工具找到与该主题相关的关键词或词组。很多索引和摘要包括按字母顺序排列的专题列表，包括交叉引用来指导你使用必要的关键词。另外，索引工具经常附有称为“词典”的册子，来指导你在使用特定的资源搜索文献时应使用哪些合适的关键词。光碟数据库则附有在线专题列表或词典引用。

相关心理学索引

《心理学摘要》（*Psychological Abstracts*, PA）和计算机版本的PsycLIT索引的书籍章节和杂志文章都与心理学有关。相关的工具是《心理学索引关键词词典》（*Thesaurus of Psychological Index Terms*, Walker, 1994）。虽然卡片或在线目录会给你提供关于某主题的详尽书籍，但还有很多书，包括由多人撰写的关于各方面主题的章节，是不被卡片和在线目录收录的。PA和PsycLIT索引是心理学相关研究的主要索引工具。

如上所述，很多心理主题可以从不同的取向加以研究。一个主要的重叠研究领域是社会学。这一领域的主要索引工具是《社会学摘要》（*Sociological Abstracts*，电子版称为社会学档案）和相关的工具《社会学搜索关键词词典》（*Thesaurus of Sociological Indexing*,

Booth, 1992)。另一个既有印刷品又有在线格式的是《社会科学索引》(*Social Science Index*)。这是一个简单的专题题目的字母顺序列表,每一专题附有某一时间段内该主题的文章引文。虽然没有专门的词典,但电子版本也含有一个“浏览器”,来指导你使用适合你的主题的专门术语。

另一个相关的领域是教育,其中最广泛使用的电子资源是ERIC(教育资源资料中心, Educational Resources Information Center)。两个相应的印刷品是《教育学杂志当前索引》(*Current Index to Journals in Education*, CIJE)和《教育资源》(*Resources in Education*, RIE)。CIJE包含杂志文章引文;RIE提供其他格式的资料,例如研究报告和会议文集,这些在其他资料中是很难找到的。这些资源也附有相关工具,如《ERIC搜索词词典》(*Thesaurus of ERIC Descriptors*, 1980),这对于在搜索中确定最合适的关键词是非常有用的。另一个教育学索引工具是《教育学索引》(*Education Index*),也有印刷品和电子版两种形式。

主要的医药学索引工具是《医药学索引》(*Index Medicus*),是电子数据库MedLine(主要面向临床的)的一部分。这些资源使用由国家医学图书馆出版的标准专题题目列表。对于与工业和组织管理学相关的主题,其主要的工具是《商业期刊索引》(*Business Periodicals Index*)——印刷品和电子版都是这个标题。

进行搜索的另一个方法是通过引文索引,比如涵盖多学科的《社会科学引文索引》(*Social Sciences Citation Index*, SSCI)以及科学引文索引(*Science Citation*, SCI)。引文索引通过提供使用同样的参考文献的文章列表帮助扩展你的搜索范围。例如,如果你通过上述提到的索引找到可能与你的研究相关的文章,你可以用合适的引文索引确认一下。如果这些文章已经被收录,你还可以找到引用这些文章的一系列文章列表。这些附加的文章可能包括与你的主题相关的材料。上面提到的引文索引都有印刷品和电子版两种形式。

当然,很多其他的搜索方法在心理学领域的文献搜索中都很有用,这取决于你采用的视角。其中一个非常有价值的资源就是:图书馆管理员。他们能告诉你图书馆订阅了哪种类型的索引工具(以及是哪种形式),对你的搜索提供其他可选择的资源,并且就如何应用这些资源提供帮助。

文献和政府文件

系统而又全面地使用上述提到的各种搜索技术,会给你提供许多关于你的主题的出版文献。但是大部分研究和文献不会被这些搜索过程覆盖,因为它们不是没有出版,就是归个人所有,或者是出版在不知名的资料中,这就是“临时文献”。并且如果你做的是穷尽搜索,找到这种文献也是重要的。这样,除了上述提到的几个步骤,如果你想找到这种“在逃文献”就应该阅读下面的内容(Rosnow & Rosnow, 1995):

1. 联系这一领域的著名作者，并向他们索要曾写过的有关你的主题的任何文章的复印件——出版过的或未出版的。还可以请他们就你的主题建议一些其他资源。做这种要求要慎重，如果你向他们提供了你目前的参考书目列表，那么可能的话，他们只能是做一些补充（而不是制作全面的列表）。
2. 使用“国家技术信息服务（询问你的图书管理员）”。这是一个关于全部研究摘要的数据库，包括会议文集和论文。
3. 使用互联网。虽然互联网上的冲浪运动每天都在变化，但你只需要很少的指导，就可以使用关键词进行搜索而得到各种类型的资料，包括那些可以给你提供更多建议的人的名字。如果你可以找到有关你的主题的电子讨论组，你可以参加讨论并获得贴在网上的文档材料。
4. 搜索有关你的主题和政府文件。虽然政府文件通常都是出版的，但有时却难以找到。政府出版物提供了大量的资料来源。政府出版办公室（GPO）是世界最大的出版商，每年出版各种可能主题的上千种文件。搜索政府文件的主要工具是《政府文件月份目录》（*Monthly Catalog of Government Documents*），用于搜索期刊的是《美国政府期刊索引》（*Index to United States Government Periodicals*），还有电子版的《政府文件索引》（*Government Document Index*）。这些工具对那些很难找到资料的研究主题特别有帮助。

组 织 文 献

我们希望 you 开始搜索文献前通读这一章，因为你需要知道组织文献的一些技巧。我们知道，有些人在搜索过程中，自以为花费很少的钱来复印每一篇文章，最后却超支，并且只阅读了其中的1/3。其实你可以在图书馆获得大量所需的材料，并节省很多时间和经费。在这一部分，我们将指导你如何做笔记、书写提纲和综述。

做笔记

如果你要做元分析，就需要记录你读过的每个研究的特定信息（见本卷Cooper & Dorr撰写的第18章）。对于大部分其他类型的文献综述，重要而有趣的观点常常在各个材料中有不同的论述。某些情况下，你可能想得到特定的引文。有时候，你想记下支持某个假设的原因。如果你综述的是经验性研究，那么可能需要记录一般的研究发现。我们将分别简短地介绍一下。无论如何，实践证明，“注解卡片”是记录研究资料的最有效和有序的方法。

注解卡片

一般来说,有两种用注解卡片记录材料的方法。第一种方法是“单观点”方法,即在每张卡片上只记录一个材料,例如,一个单独的引文(注意记下页数)、一个定义、一个统计发现以及一个理论或假设的简短解释等。

“文章摘要”法是把一个研究的总结记录在一个卡片上。这可能不如你使用“单观点”卡片记录的材料多,但是需要处理的卡片数量比较少,并且当你再次研究你的卡片时,会对该研究获得一个整体的印象。这种方法类似于摘要。我们发现,用你自己的话来总结文章比你单纯地拷贝作者的摘要要好。

使用计算机做笔记

除了用卡片做笔记,你还可以使用各种软件,将所需的材料记录在计算机上。用计算机做笔记的优点是,你可以使用各种整理方法和搜索策略来组织你的笔记。例如,如果你在开始做笔记前对所记录的内容有些想法,那么,你可以制作一个数据库,用来搜索关键词、作者、出版年限或其他资料。换句话说,你可以制作一个类似PsycLIT的书目程序,但只供你自己使用。使用计算机做笔记的另一个优点就是,你可以经常从在线的或光盘格式的索引中下载书目资料到你的个人数据库中,但是不要让这个过程代替你阅读整篇的原文。

要记录的关键信息

不管你用卡片还是计算机做笔记,至少你应该记录下列资料:(a)完整的引文(按美国心理学会标准);(b)研究目的;(c)中心假设;(d)取样类型(大小、统计特点、取样过程);(e)分析类型;(f)主要发现(如果提供了效应大小,你就可记录下来);(g)简短结论陈述;(h)(任选)参考书目(如果这是个寻找额外材料的好来源,做这一项记录是有益的)。

书写提纲

很多学生和专业人士都担心书写提纲。一些人觉得,他们的想法没有组织得足够好或者写纲要浪费时间,因为写论文时往往会改变原来的计划。通常,当你发现难以构建一个提纲时,说明你的想法还很模糊。多读一些文献,头脑中的提纲就逐渐形成。提纲对你的帮助非常大,几乎所有的好报告都附有一个提纲。幸运的是,这项工作在你的最终作品完成之前或之后做都可以(Parrott, 1994)。

如果你采取事先计划的方法,你的笔记就应按照主题组织起来的。如果你的笔记是在卡片上,就会有一堆卡片强调一个特定的主题。如果你的笔记在计算机上,则可以输

入一个代码或关键词来确定相关的主题领域。如果可能的话,尝试用文字处理软件来书写提纲。很多这类软件都有提纲格式,可以帮助你完成提纲。重要的主题间和次主题间要保持平行;在提纲中处于同一水平的观点(如Ⅰ、Ⅱ、Ⅲ)要具有同样的特征性并用同样的语法格式书写。与之类似的是,每一主题下的次主题(如A、B、C、D)也应是平行的。在书写提纲中花费越多的心思,书写报告就越容易。

如果完成报告后再写提纲,你的想法会容易产生,因为提纲结构在写的过程中也会呈现。一些人发现,做一个有计划的提纲破坏了他们的创造性,以至于他们不想再尝试做这样的事情。但是你可以在完成初稿后,为论文做一个提纲。用这种方法,你可以确认一下你的想法是否被清楚地阐述,你的主题是否处于平行的关系,从而作相应的修改。

写作的一般技巧

把文章写好并不难,但是需要练习和掌握一定的规则、惯例和一些技巧。在本章的最后一部分,我们将提供给你书写文献综述的一些技巧。要时时铭记的是,这些并不是条条框框。你或许会发现,一个不同于本文推荐的写作风格更适合你,但仍要记住的是,科学写作比其他类型的创造性写作要遵循更多的惯例,而缺少一些灵活性。

虽然文献综述的引文部分不一定词藻华丽,但应该是引人注目的。应该使读者有兴趣继续读下去。戏剧性的统计值(但绝对不是错误的或引人误入歧途的,除非是你要就此大做文章),或者有趣的引文或逸事,或者一系列问题等都可以作为文章的开头。但是,重要的是,当你第一次坐下来的时候,不要担心是否能写出一个吸引人的开头段落。这一段经常是可以后来补上的。写一个“路径地图”式的段落常常是一个好的起点,在这个段落中,可以略述你是如何组织论文以及如何发展你的观点。如果读者产生了关于你的论文的认知地图,他们将更容易跟上你的思路并理解你的论述。

使用标题组织你的论文,这对读者是有益的,并可以让你把想法组织得更好。文章的标题通常来源于提纲,注意它们是保持平行的。另外,在你的综述各部分间使用合理的过渡,用总结段落结束重要的部分来提醒读者注意你已论述的特殊观点。

总体上来说,你的文献综述要遵循从宽泛的概念或主题到具体的概念或主题这一规律。从对一个问题的论述开始,讨论和批判性地分析已有的一般理论,然后总结关于这一主题的相关研究。初学者经常犯一个错误,对于每一个提到的文章给予相同的分量。例如,在综述中把每个研究用一个段落来论述。这读起来像一个摘要——太乏味了!一个有用的技巧是,不用作者的名字作为一段的开始。如果你用了,就会用整个一段来论述那个研究。特别重要的文章可能是例外,值得作更深层次的论述,但你应该尽量将类似的研究综合起来,并总结重要的观点、主要的发现或有趣的矛盾。

总 结

搜索文献和书写综述是你学习和掌握的技能。我们希望已经提供了足够的技巧和指导来帮助你学习这项任务。一旦你熟悉了我们以上论述的索引和摘要资源，做一个全面的搜索所需的时间和精力就会大大减少。我们还想强调两点：（a）从不同角度进行你的搜索——不要只依赖单一的搜索方法（如，PsycLIT）；（b）向图书管理员寻求帮助。

推 荐 阅 读

我们发现下述资料非常有助于在心理学领域中搜索文献和准备论文：

Borchardt 和 Francis 的《如何在心理学中找到门路：对查找文献和研究方法的指导》（*How to Find out in Psychology: A Guide to the Literature and Methods of Research*, 1984）包括查找心理学相关文献的大量书目资料 and 提供了做研究和思考问题的良好建议。

虽然 Hunter 和 Schmidt 的《元分析方法》（*Methods of Meta-Analysis*, 1990）的重点是介绍进行元分析的先进工具，但第12章（定位、选择和评价研究）提供了查找心理学文献全面的、逐步的指导。这对于查找“在逃文献”尤其有益，它也对组织全面的文献综述提出了总体建议。

Parrott 的《如何写心理学论文》（*Write Psychology Papers*, 1994）对于心理学系的本科生很有帮助。它易于阅读，并且不仅对组织文献综述，还对如何选择主题、使用图书馆、查找参考资料、书写提纲等提出了建议。Parrott 提供了一个出色的书目。

Rosnow 和 Rosnow 的《心理学论文写作》（*Writing Papers in Psychology*, 1995）与 Parrott（1994）相似。这本书容易阅读，包括一个全面的书目，并且在使用各种参考资料方面提供了指导。

另外一些有用的百科全书是 Corsini 的《心理学百科全书》（*Encyclopedia of Psychology*, 1994）、Gegory 的《牛津思想实录》（*Oxford Companion to the Mind*, 1987）和 Ramachandran 的《人类行为百科全书》（*Encyclopedia of Human Behavior*, 1994）。

《心理学书目指导》（*Bibliographic Guide to Psychology*）是出版物目录的年度综合列表，这些出版物在过去几年都是由纽约公共图书馆的研究室来编写目录的，并得到了来自国会 MARC 磁带内容的补充。

一些有用的手册有《Steven 实验心理学手册》（*Stevens' Handbook of Experimental Psychology*, Atkinson, 1988）、《行为矫正和治疗国际手册》（*International Handbook*

of Behavior Modification and Therapy, Bellack, Hersen, & Kazdin, 1990)、《神经心理学手册》(Handbook of Neuropsychology, Boller & Grafman, 1988—1991)、《工业和组织管理心理学手册》(Handbook of Industrial and Organizational Psychology, Dunnette & Hough, 1990—1994)、《心理学国际手册》(International Handbook of Psychology, Gilgen & Gilgen, 1987)以及《认知生理心理学手册》(Cognitive Psychophysiology, Jennings & Coles, 1991)。

其他的参考资料,你可以参考美国心理学会的杂志《当代心理学》(Contemporary Psychology),这完全是关于书籍回顾的;《社会和行为科学当前内容》(Current Contents in Social and Behavioral Sciences),这里有即将出版的心理学相关杂志的内容列表;《累加书籍索引》(Cumulative Book Index),对专题文献的题目提供全面的报告;Sheehy的《参考书指导》(Guide to Reference Books, 1986);《全国协会目录:作者名单》(National Union catalog: A Cumulative Author List)提供了查找作者的书目信息;Wang的《行为科学杂志作者指导》(Guide to Journals in the Behavioral Sciences, 1989)。

最后,对怎样进行文献综述的其他指导,可见Cooper(1984)和Stewart(1984)。

参 考 文 献

- Atkinson, R. C. (Ed.). (1988). *Stevens' handbook of experimental psychology* (2nd ed.). New York: John Wiley.
- Bellack, A. S., Hersen, M., & Kazdin, A. E. (Eds.). (1990). *International handbook of behavior modification and therapy* (2nd ed.). New York: Plenum.
- Boller, F., & Grafman, J. (Eds.). (1988-1991). *Handbook of neuropsychology* (5 vols). New York: Elsevier.
- Booth, B. (1992). *Thesaurus of sociological indexing terms* (3rd ed.). San Diego: Sociological Abstracts, Inc.
- Borchardt, D. H., & Francis, R. D. (1984). *How to find out in psychology: A guide to the literature and methods of research*. New York: Pergamon.
- Cooper, H. (1984). *The integrative research review: A systematic approach*. Beverly Hills, CA: Sage.
- Corsini, R. J. (Ed.). (1994). *Encyclopedia of psychology* (2nd ed., 4 vols.). New York: John Wiley.
- Dunnette, M. D., & Hough, L. M. (Eds.). (1990-1994). *Handbook of industrial and organizational psychology* (2nd ed., 4 vols.). Palo Alto, CA: Consulting Psychologists Press.
- Gilgen, A. R., & Gilgen, C. K. (Eds.). (1987). *International handbook of psychology*. New York: Greenwood.
- Gregory, R. L. (Ed.). (1987). *The Oxford companion to the mind*. New York: Oxford University Press.
- Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis*. Newbury Park, CA: Sage.
- Jennings, J. R., & Coles, M. G. H. (Eds.). (1991). *Handbook of cognitive psychophysiology*. New York: John Wiley.
- Parrott, L., III. (1994). *How to write psychology papers*. New York: HarperCollins.
- Ramachandran, V. S. (1994). *Encyclopedia of human behavior* (4 vols.). San Diego: Academic Press.
- Rosnow, R. L., & Rosnow, M. (1995). *Writing papers in psychology* (3rd ed.). Pacific Grove, CA: Brooks/Cole.
- Sheehy, E. P. (1986). *Guide to reference books*. Chicago: American Library Association.
- Stewart, D. W. (1984). *Secondary research*. Beverly Hills, CA: Sage.
- Thesaurus of ERIC descriptors* (8th ed.). (1980). Phoenix, AZ: Oryz.
- Walker, A., Jr. (1994). *Thesaurus of psychological index terms* (7th ed.). Washington, DC: American Psychological Association.
- Wang, A. Y. (1989). *Author's guide to journals in the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.

第 4 章

阅读和评价研究文章

Kathryn C. Oleson

Robert M. Arkin

阅读研究文章与写作一样，充满了艺术和科学。本章我们将讨论评论性阅读的技巧。但是，如果你打算进一步扩展你的研究性阅读能力，我们也极力建议你读一读并且思考一下第21~24章。学会阅读研究文献和学会写作密不可分。有效的阅读和写作都是批判性思考技能的反映（Squire, 1983）。两者都与交流有关，一方面是接收信息，另一方面则是传递信息。有效地传递信息和有效地接受信息互为对称。要成为一个好的信息传递者，至关重要的是把你自己放到接受者的位置上；要成为一个好的接受者，则要想象作者试图传递什么信息。

本章的中心内容是描述一套在阅读、评述和评价研究中要用到的工具。其中的一些更像精密仪器，而不是铲子和钻子这些普通工具。只有大量的研究经验且对研究有深入的体会才能很好地使用这些“精密仪器”——如深入剖析研究设计时要用到的工具，考察研究者所用的多种统计方法时要用到的工具。另外的一些更普通的工具你则已经非常熟悉。

从我们当教师的经历中发现，对初学者来说，阅读研究文章和写作一样使他们感到焦虑和沮丧。我们的关于精密仪器和普通工具的比喻是为了帮助你那些顾虑抛开。批判性阅读有很多层次，即使没有完成“研究设计和统计分析的精密工具101”这门课程，你也可以在某些层次上做得很好。

目的和听众

实际上，至关紧要的起点在你阅读文章之前。如果先问过自己如下的一些问题，你将得到更高的效率和更好的效果：（1）我读这篇文章的目的是什么？（2）我将面对什么样的观众对文章下评语？有时候，人们阅读心理学的文章纯粹是为了娱乐；更多的时候，你会在头脑中持有某种目的，包括把你读的内容告诉别人，或传达给别人你对文章的评价。

你的目的是正式或非正式的，取决于你是阅读一项研究还是对整个文献进行调查

(图4.1)。对一个课堂作业,你可能只要从头到尾地描述一项研究就够了,可能还需要对其优缺点进行简短的评价。你的任务是在研究文章中找到你的思路,理解它要讲述的故事,提供一些批判性的分析。在另外一些情境中,你的目的是阅读并且综合一系列的文章(文献综述)。为此目的,你的第一次阅读除了学习在研究文献中的一般规律外,还有特殊的目的,如a)发现使用同一种研究方法的研究;b)找到那些揭示相同发现的研究;c)确认那些发源于同一个理论的研究(参看第1~3章,其中提供了十分有益的建议:找到研究主题,有效使用图书馆和进行文献搜索)。

	文章的数量	
	一篇或几篇	很多篇
非正式 (如教育目的——你的目的是学习;听众是同学、辅导员或同事;通常是口头陈述或对话)	* 课堂作业(如口头报告) * 背景阅读	* 文献综述 * 数据分析
正式 (如学习目的——你的目的是评价或决策;听众是作者、编辑和其他的评论者;是文字形式)	* 杂志文章回顾	* 你刚被任命为《好思想和好研究杂志》(<i>Journal of Good Ideas and Good Research</i>)的主编

图4.1 你在读文献过程中的目的和听众

最后,你可能应杂志主编(或者副主编)之邀,要拿出对提交给正式出版物的手稿的正式评价。这个评价的目的一般是:a)对主编是否应该发表提出建议;b)提供你对这篇文章的一些评语,判断其优劣——涵盖的范围从理论贡献到研究设计和统计分析(参看第24章)。

蜻蜓点水的阅读

一旦明确你的目的和听众,你就准备开始阅读了。但是在此,我们还是建议你要避免“马上就做”的冲动,而要以“蜻蜓点水式的阅读”作为开始。

阅读和写作领域的专家推荐从一本书的前言和序言开始读。首先,从这些篇幅中,你能获知作者关注什么,他们的同行和同事是哪些人,他们把研究放在哪些传统和学识的流派上。其次,纵览章节的标题可以了解一本书的主题,知道作者将怎样进行逐步阐述(Rheingold, 1994)。只有此时,你才能开始读第1章——紧跟着读最后一章!这些步骤合起来,你就可以对此书得到一个比较完整的认识。如果你的胃口被调了起来,你就应该

阅读中间章节的第一段和最后一段，浏览书后索引，最后详细阅读中间章节（Rheingold, 1994）。不过几个小时，你就会发现这本书的主要思想，熟悉作者的风格，并且弄通书的背景。最重要的是，你逐步对这本书与你的目的和兴趣的相关程度做出了评价。在每一步，你都能放弃浏览，转移到下一个学习任务上，这种策略既快速又高效。

你可以轻松地将同样的浏览技术用到实验研究文章的阅读上。这种有效的、蜻蜓点水式的浏览并不总是死板地依照时间顺序，从“引言”到“方法”到“结果”到“讨论”，你当然可以发挥自己的风格，但最好的始点是“摘要”。对那些没有时间和不愿意读下去的人来说，摘要起到了执行总监的作用，提供了背景、重点和特别的提示；引言部分的第一段提供了更为详细的背景；讨论的第一或第二段总结了所有发现，并揭示这项研究在多大程度上达到了它的目的。接下来可以看看研究假设部分（通常在引言的结束处），然后紧跟对结果部分的简要浏览，最后是对方法部分的浏览。不管什么样的顺序，为了符合特定的目的（例如文件综述或正式评论），你的步骤是可以改变的。然而，我们认为，无论目的是什么，读完摘要之后快速而有效的浏览方式是依次阅读文章每一段的第一句（通常是主题句）。一般而言，这种高度可操纵的任务只需要几分钟，却可以达到关于文章的大量信息，而且使你比较清晰地认识到更详细的阅读是否有必要。

我们强烈建议你试试这些策略，相信你会感受深刻并且欣喜若狂。作为起步，你应该找到杂志的“简短文章”[例如《人格和社会心理学报》（*Personality and Social Psychology Bulletin*）、《咨询和临床心理学杂志》（*Journal of Consulting and Clinical Psychology*），或《健康心理学》（*Health Psychology*）]，尝试几次“第一句”策略，看一下较少的时间、精力投入能获得多少信息。然后看看你能多快地掌握文章的核心，包括特定的自变量，或几个实验性文章的普遍发现。用这种方法，在心理学的研究文章中找你自己的路，是建立初学者自信心的快速而有效的方法。

评阅一篇研究文章

这里我们将对深度评阅一篇文章提出特定的建议，无论评阅的目的是针对课堂作业还是待出版的手稿。

一般来说，评价一个手稿时，针对作者的工作提出的问题就是你自己做研究和写作所要回答的问题（参看第21~23章）。正式的研究评论包括两部分：a）根据评定内容的质量做出是否值得出版的建议；b）给作者和主编作有关手稿优缺点的建议以及如何修改的忠告。这项任务的课堂版本可能会更加简要，并且可能是口头演示。当然，作者和编辑不仅仅是听众。美国心理学会出版的《出版手册》（*Publication Manual*）第四版提供了较好的完成这项任务的摘要：

- 所研究的问题重要吗？这项工作是否具有原创性和重要性？

- 所使用的工具是否具有令人满意的信度和效度？
- 对结果的测量是否与研究所关心的变量有清晰的关系？
- 研究设计是否充分地、毫无疑义地检验了假说？
- 被试是否代表了具有普遍性的总体？
- 研究者在处理被试时是否考虑了道德标准，如是否具有欺骗性？
- 这项研究是否有足够成熟或先进的程度，能够使结果的出版有意义？（pp.3-4）

不过，潜在的标准还有很多。Gottfredson（1978）确定了83条，用于编辑和评阅人在评价手稿时使用；Lindsely（1978）将评议任务缩减为12个维度。或许并不奇怪的是，统计分析显示，评阅人之间的一致性只是中度相关（Fiske & Fogg，1990）。

评阅人一致性

说评阅人不一致，并不意味着他们意见存在分歧，而是他们倾向于针对不同的问题表达适当的和准确的观点（Fiske & Fogg，1990）。Fiske 和 Fogg（1990）进行过一项对所投稿件缺点评语的调查。调查表明，在众多评语中，评价的不同方面个数从0~37个，平均数为8.6个。对你而言，如果你选择一个评价样例，作为自己努力的方向，平均水平（少于10个）可能是最有意义的（参看本章结尾处的附录4.1）。然而，这个从无到多的范围也是有意义的。主编熟悉这种评语上的不相重叠性，实际上，他们也希望这样。主编会选择不同类型的评阅人（年老的，年轻的；专家，通才；赞成的，不赞成的，冷静的），希望他们强调不同的方面。

两个评阅人是否同意一篇文章出版，他们的评价标准可能各有侧重。一项对新问题的突破性研究可能具有很高的潜在影响力，但是，在设计和分析方法上也会受到指责（Perlman & Dean，1987）。对这两种因素的权衡在很大程度上反映了读者的价值取向。主编必须权衡建议和评语，并运用个人的认知评价以取得平衡和得出结论。这个过程需要专业知识和均衡的判断，主编就是在此基础上被选出来的。你应该知道，在评阅文章方面，根本不存在一个正确答案。你的工作也不是寻找正确答案，而是力求做到最好，最公平，把你自己的技巧和对问题的认识运用到任务上。

表4.1是一个评阅人做出评阅判断的典型例子。

另一方面，评阅人通常会对手稿值得关注的方方面面做出评价，无论是正面的，还是负面。这些评价差异很大，但与手稿本身和杂志的类型有关。然而在这些评价当中，一个非常清晰的维度是，评价的范围从宽到窄。窄狭型的焦点在方法学和统计过程上——是“精密工具”部分。宽泛型更多地关注手稿的思想、研究理论的重要性、假说是否具备说服力和竞争力、这项工作是否会令人兴奋以及是否具备创新性等等。这两种类型的评价，都可以在附录4.1中找到。重要的是，大多数文章的评阅是对狭窄的程序问题以及对宽泛

表4.1 一个杂志评阅人所填表格的典型例子

好思想和好研究杂志	
1) 对手稿的总体评价 (画钩):	
_____	接受
_____	接受, 按照建议修改
_____	拒绝, 修改后欢迎重投
_____	拒绝, 原因见所附批评意见
2) 对手稿的秘密评价	
A) 从1~10, 10代表最高分, 就以下项目对手稿打分:	
_____	文献量是否足够
_____	理论重要性
_____	定量分析是否充分
_____	研究方法学的复杂性
_____	信息传达的清晰度
_____	未来被引用可能性
B) 给主编看的其他意见	

的理论重要性、文章价值、可能产生的影响所作判断的混合体。

新手评阅人常犯的错误

新手表现出的三个倾向非常有趣。第一, 他们不时地倾向于过分挑剔, 注意缺点或明显的错误——但是这些问题并不是致命的 (见附录4.1, 评阅B)。有点讽刺的是, 新手试图进行特别精细的剖析, 但剖析的艺术是把优点从缺点中分辨出来, 这需要经验。第二, 新手倾向于过分严厉。与理性的读者相比, 新手可能更像一束激光, 对缺陷集中火力, 并且把它们看得很重。有经验的评阅人知道, 所有的研究都是有缺陷的, 所以他们通常事先就原谅了那些小毛病和非致命的错误, 而着重于根本问题。最后, 新手比理性的评阅人更倾向于把评阅意见写得更长更详细。很自然, 对编辑和作者来讲, 评阅意见越长可能就越有帮助。他们读评阅时能学到更多的东西。因为缺乏自信, 新手的评论经常会覆盖面过宽, 因为他们希望其中的一些意见能够击中要害。当这种形式极端化, 伴随着尖锐的和狭隘的倾向, 就会导致一系列问题, 甚至可能会伤害作者的感情, 而不是提供建设性的批评意见或给编辑提供信息。

支持作者是对研究文章进行评阅的重要目的。作者投入到研究和完成手稿的繁重工作应该得到肯定; 作者冒险把他的工作置于评价之下, 应该得到奖励而不是惩罚。礼节是评阅中最重要品质。

没有研究是完美的。你要找出缺点, 但更重要的是, 你要把你的评价放在缺点是否致命这一点上。如果你牢牢抓住了文章的主要目的、文章要阐述的概念问题、所用的方法、结论的清晰度以及所作结论来自事实的程度, 你就能够理解这篇文章的缺点是致命

的还是泛泛的小问题。概念性的缺点是最致命的。如果研究的问题或理论假设不符合逻辑或有缺陷，即使再光辉的实验方法也徒劳无功。只有当影响文章的结论时，实验方法上的错误才是致命的。进一步讲，就像上面所说，寻找并且指出文章的优秀之处像寻找错误一样重要。

关于如何评估论文各部分的更多细节

现在我们对《出版手册》摘要进行更详细的讨论，对研究论文的每一个部分（引言、方法、结果和讨论/结论）进行评价。需要指出的是，我们两位都是社会心理学家，所以我认为重要的东西在其他心理学领域中未必十分重要。对每一部分我们都会列出你必须考虑的重要条目，但是请注意，每一个部分应该在对文章整体框架和目的的评估之下进行。就像我们在今后会上反复指出的：“全貌”最为重要。

摘要和引言

此处需要注意的是：

- 文章的目的；
- 研究问题的重要性；
- 在概念层次上的研究思想。

评价什么/怎么评价 带着理解“为什么要做这项研究”的目的去阅读摘要和介绍。你要尽力去理解“全貌”，问自己：“这项研究、这篇文章的目的是什么？”“假说是什么？”“问题阐述得清楚吗？”

考虑所问问题的重要性。显著吗？如果是显著的，那么它会改变我们对人类行为的思考方式吗？还要考虑到问题与先前研究的关系是怎样的。研究者是否准确而有效地继承了此领域中过去的研究？研究者是否熟悉过去的工作？描述是否完备？当前的工作是改变了还是增加了我们对此问题的理解？是否增加了我们的一般知识，是否会引导其他研究者做更多的研究？阅读了对前人研究的综述之后，考虑你会做出什么样的预言。仔细地检查这些预言，既能帮助你评价作者的假说，又使你考虑所发现事实的其他可能解释。

小测试 当你读完一篇文章，你应该能够在两三句之内描述并且评价它的中心思想。你应该理解文章的概念逻辑，你也应该能够告知他人这篇文章是否成功地阐述了要讲的问题。抓住每一个机会尝试这个小测试。这是对你能力的检测，检验你是否能自如地从“接受者”模式转变到“传递者”模式。

方法部分

需要注意：

- 内在效度；
- 结构效度；
- 外在效度。

你要评价什么 阅读方法部分还是要注意将“全貌”放在脑中，考虑作者试图在概念水平上检测什么思想，需要什么步骤来实现这些思想，研究者是否成功地将他们的概念变量转化到具体的实验操作上。通过仔细地评价方法，你要决定这项研究是否有效（Aronson, Brewer, & Carlsmith, 1985; Aronson, Ellsworth, Carlsmith & Gonzales, 1990; Cook & Campbell, 1979）。考虑研究者是否达到了他们想要达到的目的，你尤其要评价三种类型的效度：内在的，结构的，外在的（更详细的讨论参见 Cook & Campbell, 1979）。

在评价内在效度的时候，你要检查是自变量还是其他变量导致了因变量的效应。在研究程序中，是否存在某些条件，它们伴随着自变量，可能导致或者至少影响到结果。评价内在效度要考虑两个主要因素：一是被试是否被随机安排到各项条件中；二是实验中是否包含恰当的控制组和对照组。如果你认为实验具备结构效度，那么，你就判定研究者实际上对他们提出的假设做了概念水平的检验。研究者是否抓住了概念并使他们的假说具体化、操作化？

最后，如果这项研究具备外部效度，研究发现就可以被推广到实验者想要应用的人群和情境当中，这些正是实验者想要应用的。有了外在效度，结果就可以应用到其他人群和情境当中。

在评价这些方法的时候，你也应该考虑，研究者是否人道地对待了他们的被试。

怎样评价 首先，全面考虑此项研究的实验设计。研究者包含了什么实验组，是否足以检验研究的问题？被试是否被随机分配到实验情景当中，是否有合适的控制组或者对照组？在常识和为人熟知的研究设计的基础上，实验是否允许研究者清楚地、毫无异议地检验他们的假说？

下一步，考虑研究所产生的实验情境。试着从被试的角度考虑：被试在这项研究中的感受是怎样的？程序是否有效？是否试图猜测实验假设？如果是这样，是否有针对猜测的测试？还有，以实验者的观点来看待实验情境，实验者是否控制了产生偏差的实验步骤？是否知道实验条件？在考虑实验情境时，要时刻注意问题或偏见对研究效度的影响，这些效度包括内在的、结构的和外在的。也就是说，是否这些问题或者偏见而不是自变量影响

了测量和感兴趣的因变量？还有，各种各样的问题是否影响了概念变量，使它以一种有意义的方式被操作化？结果是否可以被推广到其他人群当中去？

再下一步，为了直接评价结构效度，检验实验者实现概念变量的方式。实验者怎样操作他们的假设变量？这些操作是否合适？你是否从这篇文章中得到充分的信息来理解变量是如何被控制的？是否有独立的证据说明操作的实现是合理的？你是否认为被控制的变量准确地抓住了感兴趣的现象？为什么是或不是？

如果你认为操作是不合理的，考虑一下是否有更好的方式来控制变量：考虑到现有的研究，让研究者以最好的方式掌握变量是否可能？注意，研究者正在对十分有趣的新论题进行探索性的研究，他们可能缺少操作变量的理论基础。评价的标准应该是：研究者使用标准的范式时，与所使用新的、自行设计的策略相比，对前者的评价应该更加苛刻。

再下一步，考虑因变量的测量方法。运用这些操纵，你是否相信这些测量方法准确地表达了感兴趣的现象？为什么是或者不是？测量方法具有信度和效度吗？

考虑被试的取样问题。取样大小是多少（N）？采样的人口统计学（性别、种族）特征如何？人口统计学特征是否适合这项研究？问这些问题时，你要考虑这项研究是否具备外部效度，结果是否能推广到其他情境中？如果是，那么是什么情景？

最后，评价研究者对待被试者是否人道。例如，如果使用了欺骗方法，被试是否已经知情？研究者是否为被试着想，以比较人道的和敏感的方式对待被试？

结果部分

你应该考虑：

- 统计结论效度。

你要评价什么/怎么评价 在结果部分，你要评价统计结果的效度，或者从数据得出的结论的效度（参见Cook & Campbell, 1979）。第一，你应该再次考虑研究样本的大小。注意被试的数量影响到发现结果的容易程度（第17章详细论述）。小样本很难发现一个结果，而对大样本来说则容易得多。样本大小是否大到足以发现结果？样本过大以至于统计学上虽然显著，但实用显著性可能很小。通过检查样本的大小，你可以考虑，此项研究在检测自变量和因变量之间的关系时，是否足够灵敏。

下一个，考虑样本的统计检验。分析是否有效和合适？如果可能，评价它们是否符合统计学的必要的前提假设，它们是否使用了合适的误差项。如果统计检验不合适，那么，你应该考虑它们会如何影响来自数据的结论。

最后，考虑结果报告。数据是否被清楚地展示？研究者是否忽略了部分数据？他们是否去除了一些被试？如果是，为什么？他们的推理恰当吗？如果推理看起来恰当，文章还是应该报告如果包含去除的被试，结果会是什么样子。结果是否强有力——通过p值

和绝对值来表现（参见第15、16、17章）？测量方法是否可信？

小测试 结果部分经常是最难评价的部分，尤其是对一个新手来说。在评阅文章这部分的时候，我们已经确定很多事情需要注意，但是在我们确定的东西超出了你的知识和经验基础时，你应该寻找更多的信息，而不是试图虚弱无力地或者完全错误地来对付这个问题。在正式的评语中，当你缺乏专业经验或者感到无力来讲述一件事，要说出来！你的看法的正确与否要考虑从你的背景得到判断。如果你对某种东西不能确定，你要么明确承认，要么就跳过它！你可能失去了指出明显或微妙观点的机会，但犯忽视的错误通常要比超出你的知识和经验范围而做一个错误的评论要好得多。相似的是，在一个班级的演示上，在他人试图指出错误之前首先承认你自己的局限是明智的。

讨论/结论部分

应该考虑：

- 结论效度

你该评价什么/怎样评价 在讨论部分，你要评价研究者是否发现了显著的结果，并且恰当地阐述这些结果。就像文章的其他部分一样，要按照“全貌”的方式评价结论。考虑到论文的目的、要阐述的问题、使用的方法，这些发现是否能够自然地随之而来？是否作者将其发现回溯到开始时的理论框架？作者是否考虑到数据的其他解释？数据怎样与他们的理论契合？数据怎样与其他文章中的冲突数据契合？如果数据仅仅与作者的理论契合、不与其他理论契合，这种情形是最好的。他们是否讨论了其数据的局限性，如可能的人为影响和问题？

总 结

如果你的目标是成为一个心理学家，我们建议你抓住每一个机会来努力提高你的研究评价技巧。通过成为他人研究的很好用户，你将能够更好地来发展你自己的研究项目。本章我们的目的是提供给你一系列的工具，并应用到阅读、评阅和评价研究上。我们希望使用这些工具，扩展它们，并最后形成你自己的研究风格或方法。随着你的批评技巧的扩展和精炼，你会看到你作为一个理论家、方法学家、作者而不断成长。总之，沉浸到评价现有研究中去会为严肃思考重要的新问题打下基础。而且，阅读、评阅和评价研究的批评技巧的增加，也为有效地设计和完成你自己的研究思想提供了基础。

推 荐 阅 读

在引文中,我们引用了一些指南(Maher, 1978; Schwab, 1985)。其中, Maher (1978), 一个重要杂志的编辑, 提供了一个“读者的、作者的、评阅者评价研究论文的指南”; 他的标题提醒我们, 对学习阅读和学会写作之间的密切联系, 要进行开放性的观察。沿着这条脉络, 我们也推荐你APA (1994) 的《出版手册》。这是对作者和对评阅者同样有用的手册。你作为作者的进步与你评阅文献的技能是成比例并行增长的。以我们的经验, 因果关系的箭头是双向的。我们也推荐各种各样的有关评价研究的读物, 包括Aronson et al. (1990)、Cook和Campbell (1979)、Campbell和Stanley (1966)。

参 考 文 献

- American Psychological Association. (1994). *Publication manual of the American Psychological Association* (4th ed.). Washington, DC: Author.
- Aronson, E., Brewer, M., & Carlsmith, J. M. (1985). Experimentation in social psychology. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (3rd ed., pp. 441-486). New York: Random House.
- Aronson, E., Ellsworth, P. C., Carlsmith, J. M., & Gonzales, M. H. (1990). *Methods of research in social psychology* (2nd ed.). New York: McGraw-Hill.
- Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin.
- Fiske, D. W., & Fogg, L. (1990). But the reviewers are making different criticisms of my paper! Diversity and uniqueness in reviewer comments. *American Psychologist*, 45, 591-598.
- Gottfredson, S. D. (1978). Evaluating psychological research reports: Dimensions, reliability, and correlates of quality judgments. *American Psychologist*, 33, 920-934.
- Lindzey, D. (1978). *The scientific publication system in social science: A study of the operation of leading professional journals in psychology, sociology, and social work*. San Francisco: Jossey-Bass.
- Maher, B. A. (1978). A reader's, writer's, and reviewer's guide to assessing research reports in clinical psychology. *Journal of Consulting and Clinical Psychology*, 46, 835-838.
- Perlman, D., & Dean, E. (1987). The wisdom of Solomon: Avoiding bias in the publication review process. In D. N. Jackson & J. P. Rushton (Eds.), *Scientific excellence* (pp. 204-221). Newbury Park, CA: Sage.
- Rheingold, H. (1994). *The psychologist's guide to an academic career*. Washington, DC: American Psychological Association.
- Schwab, D. P. (1985). Reviewing empirically based manuscripts: Perspectives on process. In L. L. Cummings & P. J. Frost (Eds.), *Publishing in the organizational sciences* (pp. 171-181). Homewood, IL: Richard D. Irwin.
- Squire, J. R. (1983). Composing and comprehending: Two sides of the same basic process. *Language Arts*, 60, 581-589.

附录4.1 评阅注释范例

这个附录介绍了两个简短的、说明性的评阅意见。这两篇意见都是对假想作者K.C.Oleson和R.M.Arkin写的同一篇假说手稿的评论。

第一篇举的是一个相当公平、有益的评阅例子——那种编辑愿意接收的典型类型。第二篇评阅例证了那些评阅新手、没有动力或粗心的评阅人以及恶意的评阅人常犯的错误。当编辑们收到这样的评阅时会十分担心，一些编辑具有很强的道德感，他们不会把此类评阅传递给作者。有时编辑们会寻找另外一些评阅人的评论以替代或弥补这些劣质的或没用的评阅。无论如何，出版人（包括美国心理协会）要求编辑们明确观点，指出哪些评阅人的报告他们同意而哪些不同意。多数编辑将结果写信告知作者（接收或是拒绝信），清楚地说明一些他们自己的观点，简要地总结评阅人集体的智慧。

我们知道，这两篇典型的评论并不能详尽地列举质量较好的或是无用的、挑剔的、负面的评论的特征。它们仅仅作为一个范例，让你开始做出有力的文章评论，表现出你的技巧和风格。不要为内容而读它们，而要为了解它们的格式或风格去读，这样的快速阅读会更加有益。

下面是你能在这两篇评论中发现的一些显著特质。

A

1. 评阅从对研究的总结开始。这个总结说明评阅人理解了这项研究，而这帮助评阅人、编辑和作者抓住了他们正在讨论的实质。
2. 接着，这个评阅评估了手稿在一般概念问题上的处理，讨论了它的优点和不足。它关注数据的另外解释，建议如何解决这些问题。
3. 这个评阅还列出了一些更小的问题，比如可能影响结果的方法学问题。
4. 最后，评阅以对手稿总的评价而结束，并将这个工作放在更广阔的背景之上。
5. 贯穿其中的是评阅人提出的各种建设性意见，包括改善实验设计、改善和阐明固定变量以及指出未来的研究途径。
6. 这个评阅是批评性的，也是建设性的，总的来讲是公平的。

B

评阅人B的评语既表现“忽视”错误（如它缺少我们在评阅人A的评论中看出的积极一面），又反映出了许多自己的错误。评阅B是十分恶意的、狭隘的和挑剔的。这个评阅更多的是一大堆牢骚，而不是有建设性和支持性的评语。有毫无根据的断言（其中一些可以找到根据，其他的则没有），评阅人的批评可能已远远地超出了他或她的专业特长。

好思想和好研究杂志

评阅人的评语

“对结果的自我归因中有自助偏见吗？”

Kathryn C.Oleson & Robert M.Arkin

Ms. #96-777

日期：1996年1月1日

评阅人：A

这个有趣的研究目的在于调查是否有如下的倾向：与失败的结果相比，人们更愿意将成功的结果归于个人的责任。就像作者说的，这反映了一种自助。该报告好像系统说明这个现象已迈出了合理的和潜在重要的第一步，这一步可能有理论和实践两方面的启示。研究包括两种条件：被试知道他们在假想的“社会敏感测试”（反馈是“虚假”的）上是成功了（占85%）还是失败了（占15%）。被试在成功的条件下比在失败的条件下更倾向于将责任归因于自己（他们的能力和努力）。被试在失败时比在成功时更愿意列举一些诸如任务困难和运气不好的解释。

这项研究好像是对自助偏见现象的第一次有用的证明，然而单单这项研究不能准确地探究结果背后的心理过程。因此，这个对成功和失败结果的自我归因差异，虽然是有趣的，具有启发性和新颖性，似乎应该发展，但令人失望的是，初始的发现并没有引起更加深入的对心理过程的探查。以下观点来自于这个考虑。

第一，目前的发现没有阐明方向性：这个结果应归于成功责任的强化，还是失败责任的减弱，抑或两者都有？没有中性组或控制组，这个现象的精确来源并不清楚。控制组可能反映“适度的成功”（例如60%）或对结果根本没有任何信息——两个组都可以包括在实验设计中。

缺乏控制条件引发了困扰我的另一个理论问题。虽然作者对归因不同的“动机”（即自助偏见）的解释似乎是可以成立的，但有其他的解释也同样可以说得通。例如，成功一般与一个人的努力和个人的能力有关；因此，被试有可能对成功的原因仅是根据这种相关做出猜测；这个更像是基于信息加工而非动机的解释。Kelley和其他人对这个共变的观点有很多阐述，Cunningham对这一点有一个很好的研究。另外，这些受到严密观察的被试是否有可能仅在此种情形下如此表达（而在私下表现并不会像在回答问卷时那样）？还有另外的替代解释。我想如果能提供一些有关“过程”的证据，对发现的解释更加清楚和引人信服，这个最初的“自欺欺人”证明会更加有力。这个研究仅是“初步调查”。

除了这些一般性问题之外，我还要提出三个小的问题：

1. 归因测试只关注于一个维度——从自我归因到任务归因，这样可能有问题。这个策略妨碍了我们搞清楚结果是大部分在或唯独在自我归因范围还是任务归因范围（但不在两方面）。我建议以后将它分成两个独立的维度去测量。
2. 实验者使用的“虚假反馈”让我担心，这有两个原因：（1）没有对实验者在指导语上使用盲

法,描述(要求有更多良好的操作能力)给了形成实验者偏见的机会。(2)从手稿中很难得出,被试在实验后是否得到全面、准确、详细的解说,这引起我对道德方面的担忧。

3. 讨论部分应该被修剪一下。它太散漫,几个对此工作及其启示进行解释的机会被疏忽掉了。这些本来能被包括进去,甚至占用更少的篇幅。这些启示包括以下部分:(1)人与人之间的对抗能否归结于自助偏见,因为一小部分人将荣誉归为集体的成功(或失败也归于集体);(2)这种效应是否受到我们文化中盛行的“谦虚”动机的调节?如果被试的谦虚会受到奖励,他还会有自助偏见的倾向吗?(3)在自助偏见倾向上有无个体的差异?它能解释像烦躁、喜怒无常、不开心这些东西吗?

结论

总之,我认为研究的逻辑设计和表述都很好。这项研究在某些方面上确实具有开创性,还没有任何有关这方面的论文发表。我认为,研究者在其范式中抓住了主要的概念因素。总体来看,我倾向于推荐发表这篇文章。同时,我希望作者已经做好了后续的重复研究,包含我提出的控制条件以及区分不同的因变量。加上这两点,这个工作将对文献资料作出重要的补充。



好思想和好研究杂志

评阅人评语

“对结果的自我归因中有自助偏见吗？”

Kathryn C.Oleson & Robert M.Arkin

Ms. #96-777

日期：1996年1月1日

评阅人：B

这项研究好像有本质的缺陷。作者从显而易见的观点开始——有点常识就可得出——人们更愿意承担成功而非失败的责任——但作者并没有揭示出新的东西。

当我读手稿时，我发现在文章及其报告的研究中有一些明显的问题：

1. 研究中被试的数量太少。他们应该有更大的样本量。
2. 有一大堆错字，一页甚至有两个。
3. 使用的统计方法（用t检验来比较两种情况）不合适。他们应该更注意一下所用的统计方法。
4. 他们没有包括所有必要的控制组。我简直不能相信，他们居然没有包括一个没有反馈或者仅是简单描述成功的中性控制条件。没有这些控制，我们从这项单一的研究中就学不到任何东西。
5. 实验的草案和范式似乎也有缺陷。15%的成绩似乎太低了，被试难道能相信这些吗？我不相信；我想失败的比例应该更高一些，显得更合理——如40%。
6. 作者没有考虑给他们的发现作其他解释。效应背后的机制到底是什么？它一定就是“自助偏见”吗？
7. 手稿太长，需要被压缩。
8. 研究看起来不那么严格。我想某种情况下，一个实验分两种条件是可以的，但这种“简化”太过分就变为“简单”了。
9. 作者看起来好像不知道在这方面已有研究。我虽然现在想不起来，但我知道肯定有一些。
10. 测量方法似乎无效。
11. 作者在要求被试做不同的评分时，没有进行交叉平衡。我总有这样的感觉，他们没有报告所有的发现。

我认为，本研究没有任何范围或程度上的贡献，它在整个动机领域没有什么价值。这篇文章不配称为社会认知心理学，而这个领域的方向是很清晰的。所以这项研究只不过是过时的、老式的研究。我想，这样的文章不能发表，许多读者也没有兴趣读这样的文章。它没提出任何实用的问题，理论启示也没什么意思。

第二部分

设计、选择工具以及取样



第 5 章

设计一个实验研究

Bruce E. Wampold

当你通过汽车经销商去购买一辆自己理想中的汽车时，你会发现，没有一辆车能满足你所有的标准。标准当然包括可信度、实用性、舒适程度、价格、再次卖出的价值以及视觉上的效果——很显然，考虑各种条件时往往会顾此失彼。你可能希望拥有一辆沃尔沃，但从金钱上却负担不起。而且，你想要的汽车往往并不能使别人满意（如你的伴侣）。甚至在决定了购买某种型号和样式的汽车后，你可能还要选择改变它，然后进一步增加配件，比如底板的垫子、垃圾容器等。即使是最老练的买主也时常失望，也许是因为用他（她）的标准衡量不是最理想的，或者是因为偶然因素造成的（如某汽车具有一个不为人知的缺陷）。当你拥有汽车之后，也可能会发生一些事情，改变你对它的满意度，比如：新出来的车恰好具有你的车所不具备的优点，你对汽车的需求和态度改变了，或你的汽车在不断地耗损。

设计一个实验和购买汽车有着很多相似之处。为了选择合适的设计，一个人需要有：（1）设计本身应该达到的某个特定目标；（2）一些标准，用来评判这个设计是否具有足够大的、完成特定目标的可能性；（3）对设计的逻辑认识，以便可以不断地完善该设计，从而能最优化地实现目标和避免出现错误。正如没有最完美的汽车一样，也没有最完美的实验。对一个特定设计的选择包括权衡已知因素各自的利弊，往往会顾此失彼；更进一步，可能会发生一些不可预期的事件，实验者需要修改设计，甚至可能需要终止实验或设计一个不同的实验。某些研究将成为经典，而另一些可能将被遗忘。

本章描述的是包含在心理学实验设计中的一般性考虑。仅仅了解研究的规则并将它们运用在某个你感兴趣的领域，不一定能设计出理想的实验；设计实验是一个创造过程，有时需要独创性，寻找一个简洁的方法，从而发现人类行为的真理。

研究的目的

设计实验研究的指导性原则是：确保研究本身和目的之间的一致性。通俗地说，研究的目的是为了寻求知识，了解以前未知的事情。除非已经谨慎彻底地思考过、充分地表述了研究的目的以外，为研究而做的设计不可能完整。

在科学界认可的观点中，研究是为了扬弃当前理论或发展新的理论。理论和实验性研究之间的关系很复杂，深深地包含在科学、哲学的范畴之中。然而，这种关系的要义有助于理解实验设计应与理论命题之间如何联系。对理论的一种看法，它是对心理学结构之间关系的一般性陈述。为了能进行研究，以提供有关该理论的信息，检验理论的各种含义是必要的。也就是说，如果理论完备地解释了真实的情形，那么，就可以预测在特定的情景下，某个事件或者一系列的事件可能发生。例如，如果抑郁是由于环境中缺少正面强化所导致的，那么，增加正面强化应该能降低抑郁者的抑郁程度。当含义成为明确的预测或研究假设（research hypothesis）时，研究就集中到一点之上了。设计研究之后，实验性的结果不外乎分为两类：（1）实验结果与预期相符；（2）实验结果与预期不符。

如果实验结果与预测是一致的，我们对理论的信任度将增加。反之，如果实验结果与预测不一致，那么该理论就令人怀疑了。不管研究的设计有多仔细，缺陷都会存在，因此对于该理论应作出一个适当保守的结论。当研究的证据很强的时候，理论就应该被修改、推翻或者深化。

知识在重复的实验过程中积累起来的。研究人类的吸烟行为与健康关系的每个实验都是有缺陷的，但是综合所有的结果，得出吸烟导致身体疾病的结论却是必然的。一个实验存在的问题会被另外一个实验更正，即使后者仍然存在问题。没有完美无缺的实验，但是不断积累起来的结论会不断地排除其他竞争的观点。

我们应该清楚，一个人对实验结果的信任来自于实验的质量。所以，为了设计一个好的实验，我们必须认识到可能发生的问题。效度常用于评价一个实验的“良好程度”。一方面，如果实验的效度很高，人们就会相信该实验的结果，而且会仔细地考虑结果对理论的含义；另一方面，如果效度很低，那么结论将失去意义，而且对该理论的进一步修正都是愚蠢的。效度经常用“威胁”这个概念来说明。威胁是指对一个特定实验结果另外的可能解释，而不是研究者预期的解释。一个有效的实验应该没有或者只有较少的威胁。

发现一个研究主题（Leong & Pfaltzgraff，第1章），利用图书馆有效地了解一个领域中的文献（Reed & Baxter，第2章），系统地进行文献检索（Stockdale & Kenny，第3章），回顾一个个的研究论文（Oleson & Arkin，第4章），并把理论应用于研究当中（Gelso，第29章），这些论题在本书中的其他地方都有探讨。

下面我们将讨论：（1）实验设计的步骤和研究结论的推导；（2）研究设计的类

型；（3）效度的分类和对这些效度的威胁。

设计和推理过程

研究有多个阶段，在每个阶段中都要进行关键性的推理，可以考虑如下的检验假设过程。

第一，产生一个反映理论含义的研究假设。假想一个人有一个理论：与人格相兼容的环境是形成稳定人格的原因之一。该理论的一种含义可能是，外向的人选择与其他人互动的活动，而内向的人却选择独自行动的模式。

研究假设是一种关于两个或多个结构如何相关的预期。研究假设具有一般组成形式：在一个特定的条件下，一种现象会发生。很多人认为，研究假设的发展是最具创造性且最难做到的一步。显然，拥有一个领域中的主题知识是发展假设的先决条件。有些假设通过精细化来扩展知识；另一些假设则融和各个分散的主题，通过改变领域的方向而对这个领域产生深远的影响（Platt, 1964）。“检验关键假设将产生对知识的扩展、理论的完善、矛盾的澄清、治疗中积极成分的确定等。然而，不合逻辑的研究假设……不会有解决方案，因为其不能使知识聚合到一起”（Wampold, Davis, & Good, 1990, p.362）。

第二，要选择一些操作定义来表征假设中所指的构念。在一个外向/内向区分和对环境（包括社会性互动的不同程度）选择的例子中，外向/内向和对环境的选择必须在研究中操作化。研究者也许选择使用纸—笔测验，来测量外向/内向（见Ponterotto, 本书第6章，关于选择工具的讨论）。环境中的社会化过程可以这样被设计成实验：选取两项活动，它们只有社会化程度不同，然后让被试选择其一。举例来说，可以引导研究中的参与者，使他们相信主试者正在评估他们的问题解决能力。然后给予他们选择权：是参加一个解决问题的小组，还是靠自己解决。个体独自解决和参加小组就反映了对环境选择的操作化。

第三，设计实验研究时，要考虑测量在操作量之间的共变关系。也就是说，通过心理结构操作化所得到的分数之间是否具有系统相关？还是这个例子，研究者可以测量参与者的外向/内向，然后让他们选择问题解决的情境；用这样的方法，我们可以确定，与内向高分者相比，外向高分者是否更倾向于选择小组问题解决情境。也就是说，研究者设计实验时，要注意测量变量之间的共变。在实验研究中应该可以观察到，操作自变量的变化可以对因变量产生效应。

第四，确定测量的变量是否真的像期待的那样共变了。也就是说，我们要检验结果，并确定它是否反映了我们的预期。该检验通常使用统计分析来完成。图5.1反映的是设计一个研究时的一般流程。

设计好可操作的实验后，研究者按上面描述的步骤，反过来进行推理。图5.1描述的

是如何从统计分析反推到理论层次。测量的变量确实有共变关系；自变量是造成因变量随之共变的原因；自变量和因变量精确地反映了研究者所关心的结构；结果可以推广到不同的人群、环境以及时间；而且结果对理论和/或实践有重要的含义。

在说明研究效度之前，本章讨论了各种各样的研究设计类型。研究设计的主题存在很多变式，一个有效的、将研究设计概念化的方法是将它们分类，比如被动的、实验性的或准实验性的。

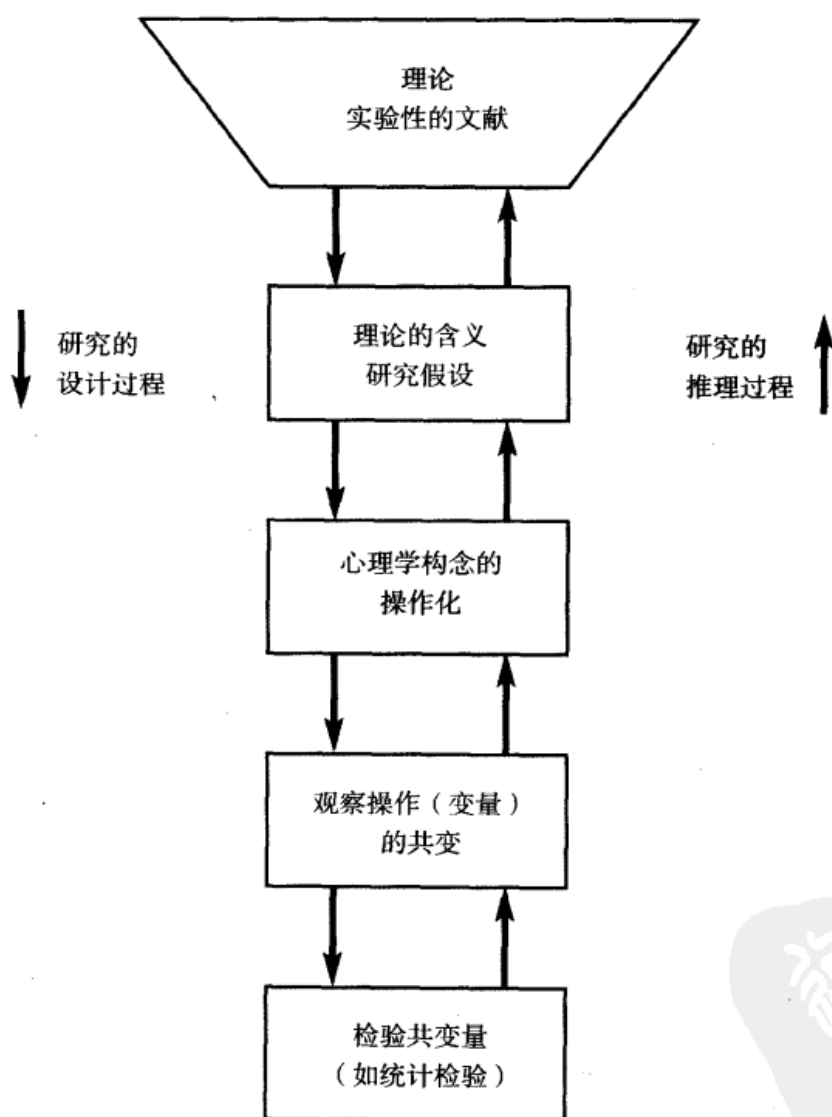


图5.1 心理学研究中设计和推理的流程

研究设计的类型

被动设计研究设计类型

被动设计是指研究者考察变量之间的共变关系，但不操作实验的某些方面而采用的设计。通常被动设计有两类：相关和组间设计。如图5.2所示，相关设计包括在一个整体中随机选择样本，在每个被试上测量两个变量，计算相关系数 r_{AB} ，然后检验这两个变量之间的真实相关性是否为零（ $H_0: \rho_{AB}=0$ ；Yaffe、Dickter & Roznowski，本书第15、16章对统计检验逻辑的讨论）。举例来说，一位研究者希望了解某群体中抑郁和焦虑之间的共变关系。他从群体中随机抽取被试，测量他们压抑和焦虑的程度（可通过压抑和焦虑问卷来完成），计算抑郁和焦虑程度之间的相关，并检验相关的显著性。如果样本的相关足够大，一个零总体相关的虚无假设就被推翻了，研究者就可以推断总体的抑郁和焦虑的共变关系。因为这是一个被动设计，研究者不可能从设计中推出这样的因果关系：抑郁引起焦虑，焦虑引起压抑，或者它们都可能是由第三个结构引起。

组间被动设计包括从两个整体中随机提取样本，如同图5.2所示。其目的是判断两个（或更多）总体的真实（i.e. 总体）平均值是否相同。样本的平均值 M_A 和 M_B 被计算出来，如果这些样本的平均值存在显著的差异，那么，将拒绝总体平均值相等（i.e., $H_0: \mu_A = \mu_B$ ）的虚无假设（可通过t检验或方差分析）。然后就可以得出结论说，总体平均值存在差异。比如，有人想测量男性和女性解析表情能力的差异。因为这是个被动设计，没有因果关系推理，所以发现的任何差异都可能是由于社会化、生物学的差异，或者动机不同而造成的。

实验设计

在实验设计中，研究者通过操纵某些变量，对各实验组进行比较，产生有关的因果关系。实验设计的目的是创造各种条件（也可称为组、处理方式或水平），这些条件仅在研究者设定的某一方面不同。这种被操纵的变量叫做自变量。图5.3中举出的就是最简单的两个实验组的设计。在总体中随机选取被试，然后再将被试随机分配到两种条件A和B当中。按照这种方法，两种条件下的任何差异都是由于偶然因素造成的。接着，研究者要求两个组的被试分别按照预先设定的、不同的方式做一些事，然后对被试在相关变量上的情况进行评定，这种变量被称为因变量，同时研究者还要计算样本平均数（ M_A 和 M_B ）；如果两组的样本平均数差异足够大，那么就认为无总体差异的虚无假设（比如， $H_0: \mu_A = \mu_B$ ）被推翻，从而得出被操纵的变量是造成差异的原因。举例来说，为了考察评估对成绩的作用，实验设计应该做到：使A组被试相信他们在某些任务中的成绩会由研究小组来评估；使B组被试相信他们的成绩会由电脑汇总，研究小组对个体情况一无所知。

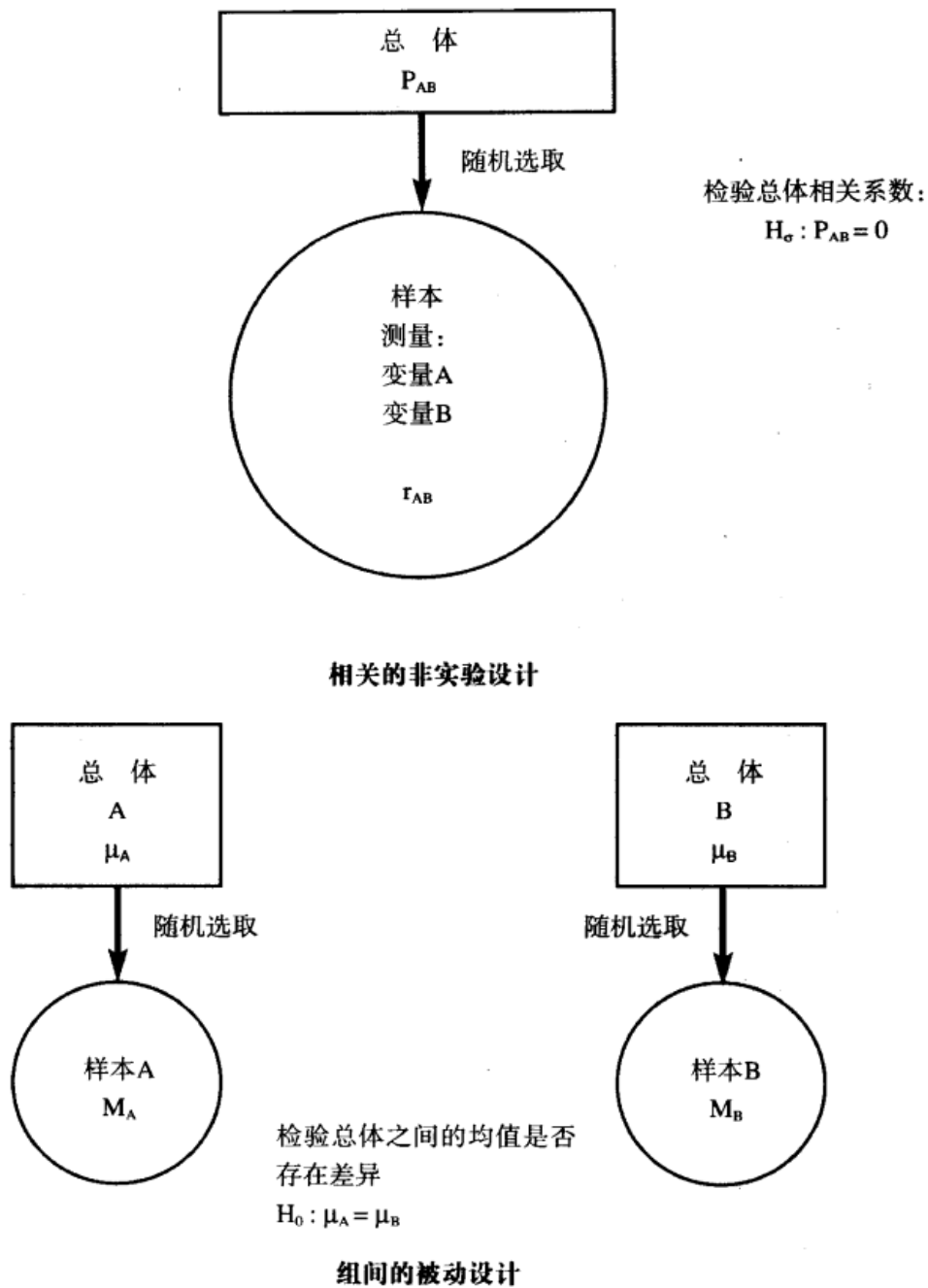


图5.2 被动设计

理论上,这两组被试仅在一方面不同,即他们的成绩会不会由心理学家来评定。这样,被试的成绩(因变量)差异可归因于评估所起的作用。

实验设计可以有很多种变式,当然也可包括多于两种的条件(比如,公众评估、个人评估和没有任何评估)。研究者可能会希望有两个自变量,每个自变量存在两个或多个水平。在上面提到的关于成绩的例子中,研究者也可能将任务的难度作为一个自变量(比如,分成两组,简单的和复杂的)。被试参与同一自变量不同水平的任务,即重复

测量的实验设计，在研究中也是经常用的。比如，在上述实验中，可要求被试既完成简单任务、又完成复杂任务。

区分被动设计和实验设计是非常有必要的。这两者之间的关键性差别在于：在实验设计中，研究者要划分各种情况，各组被试之间仅在设定的某一方面不同。实验设计和被动设计可能会考察同样的统计假设（比如， $H_0: \mu_A = \mu_B$ ），使用同样的统计工具（例如，t检验），但是两者的逻辑是完全不同的。事情也可能变得更为复杂，实验设计中也可能包含被动的方面，比如在上述实验设计中，可以包括一个非实验性的变量——性别。

准实验设计

在很多情况下，真正的实验设计是很难进行的，原因可能是伦理道德的约束、实际条件的限制、研究者为使实验情境更接近于“真实世界”而选择现场实验。接近于真正设计被称为准实验设计。从操作定义来看，凡是效度可能受到威胁的设计（在接下来的部分中会提到）就是准实验设计。有时，当研究者试图设计一项实验时，会混进各种威胁到效度的因素，因此研究就成了准实验设计。或者研究者发现不能对一个研究问题作出实验设计，就特意作了准实验设计。在上述两种情况下，设计的逻辑是十分合理的，所以可以相信实验所得出的结论。Campbell和Stanley（1966）第一次提出了准实验设计的概念；Cook和Campbell（1979）就现场实验情境而更深入地讨论了这些问题。

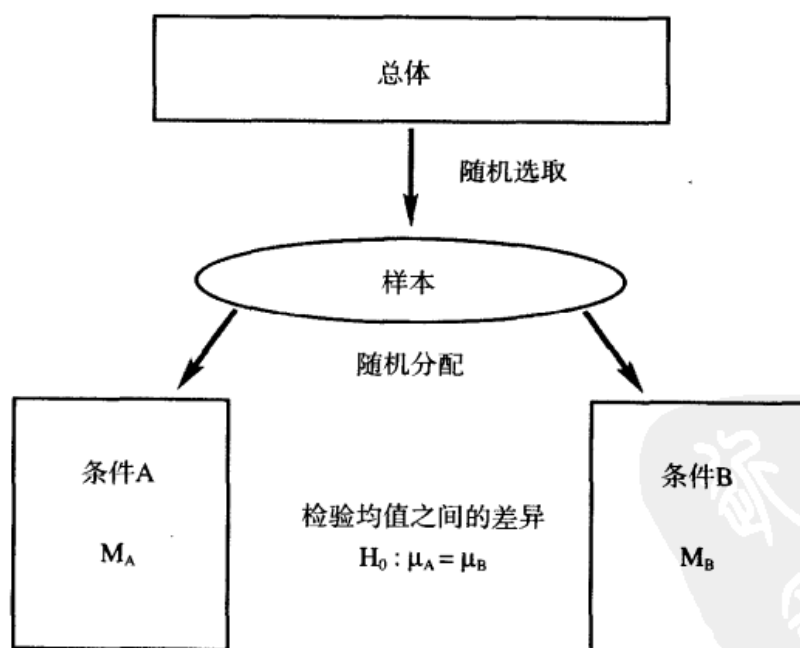


图5.3 实验设计

研 究 效 度

在研究过程中，每一步的推理都可能恰当或不恰当。如果研究的效度相当高，那么就很少有竞争性的解释，且竞争性的解释不太可信；如果效度很低，那么对已获得的结果就存在很多其他的、真实的解释。一个研究总的结论是，这些推论是否支持了假设（也就是说，所获得的结果模式与研究假设的预期相一致）或假设是否得不到支持（比如结果的模式与假设的预期不一致）。如果效度很高，那么研究者将信赖这个推理以及基于假设所获得的结论。如果有高效度，关于假设的结论就可以用来修改、推翻或发展理论命题。

Campbell和Stanley（1966）提出了两类效度：内部和外部效度。Branch和Glass（1968）扩展了外部效度的概念。考虑到现场研究中的推论的问题，Cook和Campbell（1979）后来把外部效度和内部效度分别再次分类，从而得出四种类型：统计结果效度、内部效度、假定的因果关系之间的结构效度以及外部效度。最近，Wampold等人（1990）考虑到研究的理论背景，扩展了效度的概念，提出了一些威胁假设效度的因素。下面是对各类效度的简单描述。

统计结论效度

研究中很经典的做法是：先推论各个变量之间是否真正存在共变关系。特别的是，统计检验经常被用来进行这个推论。在使用统计检验时，其虚无假设是，在一个全域中，两个（或多个）变量之间没有关系。若虚无假设被推翻，我们便可以肯定，在全域中两个变量是共变的。比如，在一个随机处理的设计中（被试随机分配到一个处理组和一个控制组），一个统计上显著的 t 检验表明，因变量的值与我们操纵的自变量的值是共变的。如果不存在真正的全域关系，那么就不太可能由几率水平造成现在这样的结果。在一个研究者没有操纵实验条件的设计中，人们可能会试图检验两个变量之间的关系，比如GRE的数学成绩和统计课上的成绩。在这些例子当中，一个统计上显著的Pearson相关系数被解释为GRE和统计课上的成绩之间线性相关足够大，以至于让研究者能够得出结论说，这并非偶然因素造成的，而是两个变量之间确实存在线性相关。

统计推论建立在概率的基础上，因此人们永远都不能确信它们是真实的。I型错误是指事实上没有真正的关系，却得出变量之间相关的结果（也就是说，拒绝了认为无关的虚无假设）。II型错误是指实际上存在相关、却判定变量之间无关的结果（也就是说，变量之间无共变）。 α 是犯I型错误的概率； β 是犯II型错误的概率，两者表明了统计结论的效度。如果一个研究设计的 α 等于0.05，那么研究者得出错误推论的可能性少于5%。Yaffee（本书第15章）更详细地讨论了统计检验的逻辑分析。

统计结论的效度受到威胁，是因为某些条件增加了错误率。比如：如果违反统计检验

的前提, 错误地推断变量共变的几率可能比实际上研究者设定的 α 更高。类似地, 如果统计效力较低, 错误推论出没有共变的可能性就高。威胁统计推论效度的因素包括较低的统计效力、违反了统计检验的前提、错误率的问题、测量的不可靠性、操作执行中的不可靠性、实验设置中随机出现的一些无关性因素以及被试的随机异质性 (Cook & Campbell, 1979)。

统计结论效度可应用到对共变的推论身上, 不管是否用到了推论性统计本身。例如, 单一被试设计常用于在一个干预和行为之间建立函数上关系 (Barlow & Hersen, 1984)。函数关系指自变量 (基线vs. 干预) 和因变量 (通常是某些行为) 的共变。这种情况下, I 型和 II 型错误仍然存在, 如人们可能错误地推论说存在真实的函数关系。

内部效度

在判定两个变量的共变 (或没有共变) 之后, 一个因果关系的归因已经形成。例如, 在一个实验设计中, 对自变量的操纵导致因变量随之改变, 操纵就是引起变化的原因。但因果关系的推断也可能是错误的。例如, 在研究中没有得到益处的参与者可能从处理组中退出, 导致他们组的平均值上升。此时退出率 (即从本研究中退出的参与者) 就成为该结果的另一种解释。内部效度表示, 在给定的变量之间作因果推断时这种推论正确的程度。

被试随机分配到不同条件下, 这些条件仅仅在一个维度上不同, 而且对各个组的处理和测评是同时进行的, 这时内部效度达到最高值。事实上, 即使在最好的条件下 (如实验室研究), 内部效度仍然受到威胁。非随机化会带来某些问题, 如执行操作之前, 组和组之间没有可比性。在研究者不对变量进行操作的设计 (如被动设计) 中, 对因果关系的推导也存在一定的危险。如在吸烟与健康的因果关系研究中, 不可能随机地把参与者分配到吸烟和不吸烟两个组中。对早期吸烟问题研究的批评有一点就在于, 吸烟者的健康不良可能是由于另外一些因素而非吸烟引起, 诸如一般性的健康行为 (吸烟者参与高风险行为)、基因特质 (有些人具有不良的健康状况)、饮食等。每一个另外解释都威胁到内部效度, 因为被试没有随机分配到各个条件中, 对被试的选择可能导致吸烟和不吸烟条件之间的系统差异。

在任何因果关系推断的背景下, 重要的一点是尽可能多地排除掉威胁内部效度的因素。内部效度的一个重要的方面是, 它与研究的设计有关, 而与使用的统计步骤无关 (Cohen, 1968)。例如, 回归分析既可以用来分析纯粹的实验设计的数据, 也可以用来分析非实验设计的数据 (Cohen, 1968; Cohen & Cohen, 1983; Kerlinger, 1985)。

正如刚才提到过的, 被试没有随机分配到不同条件的研究。在因果关系的分析上存在问题。常言道: “相关不等于因果关系”。一个简单的相关系数所表现的是两个变量之间简单的线性相关程度, 而非确认哪个是另一个的原因。设想, 已有研究表明, 抑郁的得分与社会知觉测验的得分相关, 抑郁的人比不抑郁的人更倾向于犯知觉错误, 认为

别人都不喜欢他。那么，是抑郁引起知觉错误，还是知觉错误引起了抑郁？或者是第三个变量（例如，与脑部化学变化相关的变量）引发这两个变量？近几年来，有一些人试图用统计方法检验变量因果模型（如结构方程模型，Joreskog & Sorbom, 1988；见Levy & Steelman，本书第17章中对使用高级统计检验的讨论）。然而，建立因果关系最直接的方法是实验性地操作一个变量。

对内部效度的威胁因素包括历史、成熟度、测验、工具、统计回归、选择、死亡、与选择的交互作用、对因果关系方向的模糊认识、处理方式的分散或相仿、对处理方式的弥补性平衡、被分配到不适宜条件下的被试所做的补偿性竞争以及被分配到不适宜条件下的被试对实验的反感（Cook & Campbell, 1979）。

假想因果关系和效应的结构效度

在一个研究中，被测量的变量反映的是心理学的一个结构。当测量的变量X被看作是Y的原因时，研究者想推断，操作定义为X的一个结构是操作定义为Y的一个结构的原因。结构效度是指该研究中的所有变量，包括自变量和因变量，反映的是有理论和假设所定义的理论性结构。

对自变量，实验操作需要揭示某个结构。假设一个社会心理学家相信人们更倾向于对亲近的朋友而非对陌生人进行自我倾诉，这个研究者的任务是设计出被操作变量的条件，从而使得它们仅仅在倾诉的对象（朋友vs. 陌生人）上不一样。显而易见，如果朋友是女性而陌生人是男性，条件设置就是不适当的，因为条件之间的差异可能是由性别造成，而非亲近程度。其他不适当的操作可能更为微妙。假想研究者使用一些人的照片，他们或者是朋友或者是陌生人，那么重要的一点就是确保照片除了在亲近程度上不一样以外都相匹配，这些方面包括宗教、个人魅力、年龄等等。

因变量也应该是一个结构的操作化。如果一种处理方法用来影响抑郁，那么任何对因变量的测量应该反映的是抑郁而非其他结构，如社会期望。Betz（本书第19章）讨论了如何构造一个测验来测量结构，Ponterotto（本书第6章）讨论了如何评价和选择适当的工具来操作一个结构。

如果我们不清楚一个或几个变量在研究中反映的是什么结构，这说明存在一种混淆。更明确地说，混淆是一种结构，它不能与一个假设的结构在逻辑上或统计上区分开来。在自我倾诉的研究中，如果所有的朋友都是女性而所有的陌生人都是男性，那么显然在性别和朋友/陌生人之间存在混淆。如果一个用来测量抑郁的工具事实上（在某种程度上）测量的是社会满意度，那么在抑郁和社会满意度之间就存在混淆。

很多与人的个性相关的变量经常混淆在一起。举例来说，在美国社会中，种族划分与社会经济地位相关，因此关于种族地位的归因可能更适合于社会经济地位。区分出自然产生的混淆是个棘手的问题，但通过统计方法可以部分地解决这个问题（Cohen &

Cohen, 1983)。

对结构效度的威胁包括对结构不恰当的操作化、单一操作的偏差、单一方法学的偏差、实验条件下的假设猜想、对评估的理解、研究者的期待、对结构的混淆和对结构水平的混淆、处理方法之间的交互作用、测验和处理方法之间的交互作用以及对结构之间有限的普遍性 (Cook & Campbell, 1979)。

外部效度

Cook和Campbell (1979) 把讨论的重点放在不同类型的人、环境和时间上的普遍性。外部效度的核心问题在于, 在多大程度上, 从一个研究所得出的结论能够同样推广到不同的人、环境和时间上。在不同类型的人中做推广常常最为有趣。最近的一个研究探讨了每周艰苦的锻炼与心血管疾病几率降低之间的关系。然而该研究只对男性被试做了调查。这就存在一个问题, 即男性的研究结果是否可以推广到女性被试 (并且没有得到解决)。研究中如果包括了性别的变量, 研究者就可以判明, 对男性和女性来说, 辛勤锻炼对预防疾病是否具有相同的功效。判断一个结论是否适合于不同的人群, 我们必须把不同人群作为研究设计的一部分。在因素设计中, 可以加入不同人群作为一个因素; 不同人群因素与其他一些因素之间的交互作用就说明, 结果随不同的人群而变。例如, 可以重新设计训练研究, 把性别作为一个变量, 这样就要比较4组之间的不同: 艰苦锻炼的男性, 不锻炼的男性, 艰苦锻炼的女性, 不锻炼的女性。性别和锻炼之间在心血管疾病预防能力上的交互作用说明, 锻炼对男性和女性具有不同的收益。

在不同人群之间比较某种特质, 可能的选择几乎是无限的。从政策或社会的角度来看, 重要的特质包括性别、种族、社会经济地位以及年龄。通常这里存在理论上或实际应用方面的原因, 使得我们有必要考察结论是否适用于不同的人群 (见McCready, 本书第8章, 对取样的问题和普遍的细致探讨)。

假设效度

Wampold等人 (1990) 关心方法与理论之间的关系, 探讨了假设效度。假设效度指在何种程度能够解释理论对结构之间关系所作的预测。假设效度有三个标准。首先, 研究假设 (即对结构之间关系的陈述), 对扩展知识具有关键作用。这是“谁关心”因素。研究者可以把一个研究设计和实施得非常漂亮, 但除非研究富有心理学含义, 否则几乎没有人会对结果感兴趣。

其次, 为了拥有适当的假设效度, 研究者必须以可以证伪的方式, 毫不含糊地阐述研究假设。如果假设是含糊不清的, 那么就不可能获悉结果是否与假设一致。有关“本研究的目的是探索 $x \times$ 之间的关系……”之类的陈述会导致推论问题, 因为这里没有对有关结果作出预测。任何关于两个或多个变量之间关系的相关研究将不可避免地“揭

示”某种结果模式，但是这种发现并不能用于支持或拒绝一个理论观点，因为研究者没有明确提出一个理论观点（Wampold等人，1990）。

第三，假设效度包含了研究假设和统计假设之间的密切联系。研究假设强调结构之间的关系，而这些结构被操作定义为可观测的变量，测试这些变量所得的数据要进行统计检验。在每一个统计检验底下都有一个统计假设。当采用多种统计检验方法来验证一个研究假设时，会发生这样的问题：某些检验有可能达到显著性，而同时另一些却不能。那么这些证据是推翻还是支持研究假设？理想的方法是，对于一个研究假设，只采用一个统计检验（Rownsnow & Rosenthal, 1988）。

总 结

本章探讨了研究设计中的一般注意事项。不论一个人在设计时如何仔细，仍然存在对效度的威胁，就像没有一种理想的汽车可以满足你所有的梦想一样。显然，也存在一些实现过程中的限制，特别是在研究者的时间、资源、获得被试的可能性以及伦理等方面。我们可能都渴望得到法拉利或劳斯莱斯汽车，但我们不得不满足于本田或福特。最重要的是“设计是否完成了目标”。如果设计是完备的，那么得出的结论将是可信的（也就是说，具有很好的效度），虽然这里有多种不能排除的威胁。要理解人类的行为，进一步的研究总是必要的。

推 荐 阅 读

在研究设计及其相关领域里有很多的信息来源，研究生应该向自己的导师寻求指导。Kerlinger（1985）提供了对行为研究的全面综述，它为社会科学领域研究的多个方面提供了一个相当优秀的指导。Kirk（1982）很好地陈述了实验设计的变式和每种变式下相应的统计测验。关于现场研究，Cook和Campbell（1979）探讨了在实验室以外研究的许多问题。Keren和Lewis（1993）编辑了一个册子，考察了目前在行为研究中存在的一些问题和这些研究中内在的定量方法。本章中没有探讨定性研究，可见Highlen和Finley（本书第14章）；Denzin和Lincoln（1994）编辑了一本跟定性研究范式有关的方法集。不同领域的方法学家讨论了运用于不同领域的方法，如咨询（Hepper, Kivlighan, & Wampold, 1992）和临床心理学（Kazdin, 1992）。建议研究者学会那些应用于你所感兴趣领域的最实用的设计方法。

参 考 文 献

- Barlow, D. H., & Hersen, N. (1984). *Single case experimental designs: Strategies for studying behavior change*. New York: Pergamon.
- Bracht, G. H., & Glass, G. V. (1968). The external validity of experiments. *American Educational Research Journal*, 5, 437-474.
- Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Cohen, J. (1968). Multiple regression as a general data-analytic strategy. *Psychological Bulletin*, 70, 426-443.
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavior sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin.
- Denzin, N. K., & Lincoln, Y. S. (Eds.). (1994). *Handbook of qualitative research*. Thousand Oaks, CA: Sage.
- Heppner, P. P., Kivlighan, D. M., Jr., & Wampold, B. E. (1992). *Research design in counseling*. Pacific Grove, CA: Brooks/Cole.
- Jöreskog, K., & Sörbom, D. (1988). *LISREL 7: A guide to the program and applications*. Chicago: SPSS.
- Kazdin, A. E. (1992). *Research design in clinical psychology* (2nd ed.). New York: Macmillan.
- Keren, C., & Lewis, C. (Eds.). (1993). *A handbook for data analysis in the behavioral sciences: Methodological issues*. Hillsdale, NJ: Lawrence Erlbaum.
- Kerlinger, F. N. (1985). *Foundations of behavioral research* (3rd ed.). New York: Holt, Rinehart & Winston.
- Kirk, R. E. (1982). *Experimental design: Procedures for the behavioral sciences* (2nd ed.). Monterey, CA: Brooks/Cole.
- Platt, J. R. (1964). Strong inference. *Science*, 146, 347-353.
- Rosnow, R. L., & Rosenthal, R. (1988). Focused tests of significance and effect size estimation in counseling psychology. *Journal of Counseling Psychology*, 35, 203-208.
- Wampold, B. E., Davis, B., & Good, R. H., III. (1990). Hypothesis validity of clinical research. *Journal of Consulting and Clinical Psychology*, 58, 360-367.



第 6 章

评估和选择研究工具

Joseph G. Ponterotto

为了成功完成实验研究，选择一个有效、可靠、有用的工具很有必要。心理学文献中提到过无数种工具，学生可能为它们困扰。在心理学领域中可以找到数以万计的工具（有时被称作量表或测量），几乎可以测量这个学科提出的每一种构念。其中有少数工具是精心设计、仔细地做过预实验、谨慎地修订过以及通过很多研究验证，包括清晰的常模数据和对分数的解释性指导。但是，大多数工具并没有经过仔细的考虑，也没有经过预实验。在为研究选择工具时，更多的情况是为了讨方便，而不是经过系统的评估。

本章的目的是为研究者提供一个有实际价值的向导，以便研究者能更好地选择研究工具。本章聚焦于自我报告测量法，这是心理学科中一个普遍的测量方法（请见Miller的综述，1991）。我们尤其着重于用likert式量表对观念、态度、信念进行自我报告量表的测量方法。

本章只涉及定量的工具（主要研究而使用）；对定性方案（如非结构化或者半结构化访谈、观察日志）感兴趣的学生应该多读后面的章节（第14章，Heghlen & Finley）。本章给读者在以下方面提供一些向导：查找、回顾以及评估用来测量心理构念的测量工具的性质。

查找研究工具

对工具的选择应基于对研究中涉及的心理构念已经有了充分了解（见Gelso，第29章）。阅读相关构念的文献，我们将会看到大量已发表的、有关该主题的研究，而且在这些研究中，你会找到测量构念时需要的工具。比如，一个学生对“自尊”这个构念感兴趣。一个PsycLIT或者ERIC的电脑检索会找到很多有关该主题的概念性和实验性研究。追寻着这些研究的足迹，你会发现研究中使用的各种测量自尊的工具。然后，这个学生就可以着重于文献中对“工具”和“测量”领域的描述。文章的这些部分还会提供有关

最初来源的参考资料，以及有关心理测量学详实的信息。

通常，特别是对成熟的心理学测量工具来说，人们可以通过查阅相关书籍或文献，来找到详细的、关键的综述文章。本章最后几页的附录6.1给出了一些书籍以及杂志，里面提供了心理学研究常用工具的列表、简述或者心理测量的综述。该附录里还列出了一部分主要心理工具和评估的出版人。另外，Borg和Gall（1989）还提供了很多有价值的信息资源，从中能够找到许多的测量工具。

回顾找到的工具

当研究者确定了一系列工具，下一步就是尽量减少列表中的内容。比如，假定研究者找到了10个对自尊的测量，基于已有的对工具的描述，10个当中有两三个似乎是最好的。一旦选择了这两三个工具，研究者就应该针对这些工具而不是针对工具所测量的概念进行广泛的文献检索。在这个阶段，研究者应该找到工具本身，得到使用它们的许可以及关于它的最新心理测量学信息。

获得工具

检查测量工具是工具评估过程中非常重要的步骤。通常，文献只是对工具进行描述或给出几个样例条目，因此读者不能很好地理解工具的格式和特定条目的内容。把那些实实在在的工具摆在你的面前，并以被试要填写完成的格式呈现，这对建立工具的表面效度是很重要的。我会在下面的章节里更详尽地讨论效度问题。为了获得一个实际的工具以及评分标准指导，研究者必须写信给作者或工具的发行人。版权所有者的地址通常在该工具第一次发表的文献中。另外，附录6.1中列出了多种工具的来源，包括工具的开发者或发行人的地址。

使用工具的许可

使用工具前，研究者必须写信请求对该产品的使用权。这封请求信至少应该强调研究目的、研究项目的负责人（管理人）、研究样本和研究过程以及完成测验的日期。通常在收到请求后，拥有者将邮给你使用权的许可单，如果你决定使用该测量工具，你必须填好这个单子然后寄还给他。本章节最后，附录6.2中提供了本章作者所开发的工具“Quick Discrimination Index”的使用权申请表的格式样本。

使用工具过程中的费用和义务

盈利性工具的发行人要对其客户收费。很多资深的人格、职业评估和智商测量工具需通过付费方可使用，其价格在各自的广告目录当中。但心理学平常使用的自我报告测

量工具,大多数不是由专业检测出版商提供的,而是由工具的作者控制的。在很多情况下,工具的作者允许有资格的研究者(或经过指导的学生)免费使用该工具。Quick Discrimination Index就是这样的免费工具,附录6.2中给出了许可格式样本。尽管有很多免费使用的工具,研究者在组合使用工具以及共享研究结果时仍需要承担重要的责任。

显然,学生在选择和使用心理学测量工具前,应该接受一定的训练(至少应参加过一门心理测量学课程),而且在测量过程中应该有指导人员帮助。不同的测验和工具有不同的使用和解释要求。测量者的资格水平在教育心理学和心理测量学标准里面有所涉及(美国教育研究联合会,缩写AERA,1985),在最近的Moreland, Eyde, Robertson, Primoff, and Most (1995)一文中也有所涉及。

使用测量工具时除了具有起码的能力之外,研究者还有责任对评估工具效度、信度以及实用性方面做一些贡献。考虑到所有心理测量学都有一定的局限性,工具的开发者(或发行人)有责任不断地监测工具的使用(以及潜在的误用)过程。工具的信度、效度以及实用性必须接受多次、多样本的评估,然后根据效度检验结果进行校订、修改或取消(其在未来研究中的使用)。因此,作为使用工具的条件,工具开发者经常要求工具使用者反馈研究的结果(有时候需要提供原始数据本身)。这样,工具的作者可以通过研究者的样本数据来评估工具的信度和效度,并可能把这些数据作为标准开发的一部分。

评估工具的心理测量学性质

在决定是否采纳某一工具时,信度和效度是最需要被考虑的事项。本节将对信度和效度概念进行简单的综述,并对评估心理测量学优点和局限性提出具体的建议。对信度和效度的全面综述并不在本章的范围之内。但心理学系的学生必须在选择工具之前牢牢地掌握这些概念。

心理学研究者的必要读物是《教育学和心理学测量标准》(*Standards for Educational and Psychological Testing*, AERA, 1985)以及《信度和效度的评估》(*Reliability and Validity Assessment*, Carmines & Zeller, 1979)。其次有帮助的参考来源是Sparks (1990)的《如何阅读测量手册》,它主要探讨商业和工业测验的工具选择,提出了在阅读一个测验手册以及选择工具时应注意的一些重要问题。显然,我们希望心理学系的学生事先完成一个或多个心理学测量课程,熟悉全面性的测量书籍(如Anastasi, 1982; Pedhazur & Schmelkin, 1991)。下面对信度和效度的讨论大部分来源于本段中提到的参考文献。

效 度

效度是选择研究工具时最重要的标准。如果一个工具能测量出我们真正想要测量的东西，那么它就是有效的。严格地说来，效度不是指工具本身，而是指从工具中得到的解释或推论的合适性和有用性。评估工具效度的方法很多，不同的方法得到了不同的效度类型；但这里你要清楚认识到效度是个单一的概念（AERA, 1985）。建立评估效度的传统方法有三个：内容相关、标准相关以及结构相关。可以有更为精细的效度辨别方法，如标准相关的效度可以通过同时性和预测性的方法进行评估；结构效度可以通过汇聚性和辨别性程序过程来评估。一些研究者会讨论表面效度这一概念，我们就从表面效度开始说起。

表面效度

表面效度是指研究者对测量工具内容的主观评估。通过阅读工具的实际项目，研究者可以评估及推断：“是的，这些项目看起来是测量我所设想的结构的。这些项目对我来说很合理。”相反，研究者也可能会想：“不，这些项目不是我真正想了解的，它不能反映我要研究的主题。”

与表面效度相关的另一个要考虑的问题是工具的实用性（或是否有用）。研究者想要问的问题包括：工具的长度是否适合我的研究意图（有些工具相当有用，但是它们的长度由于研究的时间限制可能不合适）？语言文字和可读性是否适合我的样本？对某些样本来说，有些项目是否可能不适合？格式和版面设计是否足够吸引人？

内容效度

表面效度是对工具的主观评估，内容效度则涉及一个系统化的评估过程，以确保工具项目内在的内容测量确实是某一个特定的领域。内容效度包括清楚地定义你的研究领域及其结构，为工具在这个领域中选择一个合适的问题样本，以及在工具的项目中清楚地加入这个合适的样本。通常研究者为了获得内容效度的证据，会请该领域内的资深专家从清晰度和领域适宜度这两方面来评估工具的每一个项目。

标准相关效度

如果工具的得分可以显示与一个和多个结果标准相关，那么可以说，该工具具有标准相关效度的证据。通常可通过两种设计来获得标准相关效度的证据。首先，当所有的数据同时收集时（即同时性效度），在工具的得分与标准的得分之间有相关性；其次，如果在使用工具收集到数据之后才进行标准数据的收集，它们之间的关系将成为预测效

度的证据。用统计方法计算工具的得分和标准的得分之间相关性后，由相关系数表示出来的指标被称为效度系数。显然，效度系数越高，则用工具来预测标准就越有效。

结构效度

结构效度关心的是两个测量结果之间的关系与理论设想出的关系是否一致。某个工具所测量的结构应符合某个概念框架，即使该框架并非完美或还没有完全开发。概念或理论框架定义结构，区分结构，并设定被测量的结构如何与其他结构相关。检验结构效度的一般程序包括运用理论来预测一个工具与相似的或不相似的结构之间的关系、采用因素分析方法探索能够最好代表工具条目的结构因素。

建立结构效度的一个常用程序是建立多重特征、多重方法矩阵（Campbell & Fiske, 1959），同时，它能同时检验汇聚效度和辨别效度。这个过程要求采用两个或多个不同的方法来评估两个或多个特征（如人的性格）。为了建立汇聚和辨别效度，研究者将预期，对同一特征的不同测量之间应有高相关（汇聚效度），而采用相似或不同的方法测量出来的不同特征之间应是低相关（辨别效度）。

一个例子可以让你清楚地认识该程序，心理障碍的两种特征分别是抑郁和焦虑。评估这些结构的两种方法是自我报告的问卷和诊断的等级量表。在这个多种特征—多重方法的矩阵结构中，人们将预期自我报告的抑郁情绪和对抑郁情绪的诊断等级之间有高相关（相同特征、不同方法），并高于在自我报告或诊断等级中抑郁和焦虑之间的相关（不同特征、相同或不同方法）。从多重特征—多重方法的矩阵中得到的发现，可以为测量工具的结构效度提供强有力的支持。这个程序在心理测量学研究中逐渐普遍化，聪明的学生可以去参阅对这个过程的早期探讨（Campbell & Fiske, 1959）和后来更为综合的探讨（Anastasi, 1982; Pedhazur & Schmelkin, 1991）。

建立结构效度在验证过程中非常重要。事实上，有些测量学专家（Pedhazur & Schmelkin, 1991）认为，所有“类型”的效度都是围绕结构效度的，也就是说，结构效度在验证中处于最核心地位。

效度小结

研究者应该选择那些经过多次重复性效度验证的工具。最重要的一点在于选择那些与你的研究意图相匹配的、有效度根据的工具。因此，如果你寻找工具去测量一个多维度的结构——如状态焦虑和特质焦虑——该工具必须有效度根据，支持状态和特质焦虑的分别使用和解释。再举一个例子，如果你在一个研究当中试图使用一个工具作为一个预测变量，那么，在进行本研究之前，该工具必须有足够的标准相关效度验证。

一个工具不需要得到所有可能“类型”效度的支持。但是你要记住，这里没有彼此毫不相干的效度：这个概念是一个统一的概念。事实上。效度“所指定的类型”（内

容、标准相关及结构)之间通常都是互相关联的。因此,对一种测量来说,当一个工具不断得到对其标准相关效度方面的证实(如抑郁和生活满意度相关,也可以预测生活满意度),则在结构效度方面也因此而得到支持(如证实抑郁这个概念的确存在,工具中的条目对这个概念进行了操作性定义)。总之,从多个工具中进行选择时,研究者必须提出的最重要的问题是:“针对我要测量的东西,这个工具是否有较高的效度?”

信 度

信度指工具得分在多大程度上避免了没有测量误差。将一个测量工具运用于某个被试所反映的是一个“得分”,不一定是一个“真正的”分数。也就是说,如果被试第二次经过该工具的测量,她或者他可能会得到不同的分数。这也许是由情绪的改变、精力水平、测量环境等造成的。在分数上如此的多变性反映的就是测量误差。被试的真实分数是他或她的“获得的”分数与测量误差的混和。当选择一个研究工具时,研究者期望测量误差最小,而获得的分数能最大程度地接近真实分数。信度系数允许研究者根据已有工具的测量能力做出明智的选择。信度系数越高,测量误差就越低,这种工具就越吸引人。

评估一个工具的信度有多种方法:重测方法、交叉平衡方法、分半法以及内部一致方法。本章中我们将概述这些基本的信度评估方法。

重测法

作为评估信度的一种方法,重测是指在短时间内对同一个样本先后测量两次。两次独立测量取得的分数之间的相关就产生了一个重测相关系数。该方法的令人担心之处在于,一个人分数的真实改变会被错误地归结为测量上的不稳定。例如,让我们假设在时间点A上,一个人完成了一个“对老年人的态度量表”;两个星期后,在时间点B上,同样的样本第二次完成这个测量。让我们假设,在这两次测量间的两星期间隔里,很多人偶然看到了“60分钟”电视节目,而这些片段恰巧反应的是老年人的生活经历和面对的挑战。这些影片可能会刺激态度的改变,因此,在时间点B上的得分会发生变化,而这使得两次测量结果的相关度降低。如果研究者没能意识到电视因素的作用,较低的相关度就可能被认为是工具的低重测信度的证据。

重测法中另外一个需要考虑的因素是:测试本身也会引起被试的改变。也就是说,完成工具的过程中会让你思考和感受你先前已有的思想和感觉,在第二次施测时,这些思想和感觉却有所变化。最后,如果重测的时间间隔非常短,可能会由于记忆的原因而获得被高估了的重测效度,因为被试会回忆起先前他或她对每个项目的回答,从而给出相似的答案。重测的时间间隔要根据研究的类型和测量的目的而变。如果工具旨在测量

相对稳定的结构（如人格、智力），那么重测要求更长的间隔。然而，测量稳定性较小的结构（如情绪）的工具需要在较短的时间间隔内来检验重测相关系数。两个星期被认为是最小重测时间间隔（Carmines & Zeller, 1979）。

交叉平衡法

类似重测法，交叉平衡法也需要对同一个总体进行两次测量。与上述方法不同的是，交叉平衡法并行实施两个版本的工具。两者当然都是为了测量同一个结构。两个版本的工具看起来在格式、项目个数等方面都很相似。建立交叉平衡信度的典型评估程序是选择工具的一种版本，两个星期之后（Nunnally, 1964）再测量另一个版本。两个版本得分之间的相关因此被解释为是对交叉平衡信度的评估。

这种方法比重测法更得到人们的信赖，那是因为平行版本的使用控制了练习效应。然而，考虑到两次表格发放之间的时间间隔，真实的得分改变（例如上面因电视节目而导致的改变）仍可能引起重测信度的下降。

分半法

与重测法和交叉平衡法不同，分半法不需对同一个被试进行两次测量。分半法只需要进行一次测量。研究者将项目分成等同的两半（如奇数和偶数项，或对项目进行随机分配），计算两者之间的相关。因为在这种情况下我们计算的是半个量表的信度，在解释整个量表的信度时，就应该进行某种统计上的修正。这个修正被称作斯皮尔曼—布朗公式。这个公式相当简单，可以在任何一本测量学或统计学书中找到。

分半法的局限性在于因分半的方法不同而得到稍微不同的信度评估。如在一个10个项目的量表中，有125种方法把所有的项目分离到两个项目组中（Carmines & Zeller, 1979）。125种分法得到的相关系数可能稍微不同。对工具信度一个更为精确的评估，应该是将这125个相关系数进行平均。这个过程奠定了使信度评估具有更好的内部一致性的基础。下面我们将对此加以探讨。

内部一致性方法

与分半法一样，内部一致性的相关系数也是只对样本做一次测量就可得到。这类相关系数最为常见的是Cronbach的 α （Cronbach, 1951）。考虑到 α 值可以被认为是对所有可能的分半系数的平均，这个方法比分半法更为人认可。

α 值被认为是对工具信度的保守评估。 α 的大小取决于项目间的相关和项目总数。单个项目间的平均相关越高量表越长（项目越多）， α 就会越大。

通过对上述4种测量信度方法透彻的分析，Carmines & Zeller（1979）对交叉平衡法和内部一致性法有强烈的好感。事实上，这些作者建议，任何多项目的量表都应该计算

alpha值 (Nunnally, 1978)。

什么是信度的满意水平

心理学系的学生以及研究者们经常提出这样一个问题：当选择一个工具时，什么是最小的、可接受的相关系数？对此没有明确的答案。最重要的考虑是，工具是否适合你的目的。如果某工具的得分会显著地改变人们的生活（如录取和选拔），那么应该要求最高水平的信度，通常在0.9以上 (Walsh & Betz, 1985)。当选择工具是为了大样本的研究目的，一个0.8 (Camines & Zeller, 1979; Sparks, 1990; Walsh & Betz, 1985) 或甚至是0.07 (Nunnally, 1978) 的alpha系数也是可以接受的。

这里需要提醒你的是，在选择一个研究工具时，信度仅仅是一个重要的评价指标。同样重要的指标是你要用的工具的效度、该工具是否容易运用到你所选择的样本身上。当所有证据被衡量过以后，研究的目的也很清楚了，一个alpha在0.70上的相关仍然是比较令人满意的。当一个研究工具用于某项研究时，我们建议，应对其信度进行重新检验，把此作为研究的一部分。

总 结

本章强调，在心理学研究中选择使用工具是研究过程中相当重要的一步。不应该以方便作为选择工具的依据，而应更多地在多种标准之间进行谨慎的评估。这些标准应该始于对要研究的心理学结构的牢固理解，对基本测量概念的了解，对一个工具的实用性、信度和效度的仔细考察为基础。如果工具并非有用（如它太冗长）、不可靠或无效，你的研究结果就会立即受到质疑。因此，我在本章的结尾提出一个实际忠告：在设计你的研究时要精明地、慎重地选择工具。

推 荐 阅 读

下面的参考文献对本章里提到的论点有深层次的讨论：Anastasia的《心理测验》 (*Psychological Testing*, 1982)；Carmines和Zeller的《效度和信度评估》 (*Reliability and Validity Assessment*, 1979)；Pedhazur和Schmelkin的《测量、设计和分析：一个综合的方法》 (*Measurement, Design, and Analysis: An Integrated Approach*, 1991)；AERA的《教育学和心理学测量标准》 (*Standards for Educational and Psychological Testing*, 1985)；Sparks的“如何阅读测验手册” (1990)。

参 考 文 献

- Anastasi, A. (1982). *Psychological testing* (5th ed.). New York: Macmillan.
- Borg, W. R., & Gall, D. G. (1989). *Educational research: An introduction* (5th ed.). New York: Longman.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validity by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Carmines E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Beverly Hills, CA: Sage.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Miller, D. C. (Ed.). (1991). *Handbook of research design and social measurement* (5th ed.). Thousand Oaks, CA: Sage.
- Moreland, K. L., Eyde, L. D., Robertson, G. J., Primoff, E. S., & Most, R. B. (1995). Assessment of test user qualifications: A research-based measurement procedure. *American Psychologist*, 50, 14-23.
- Nunnally, J. C. (1964). *Educational measurement and evaluation*. New York: McGraw-Hill.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
- Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach* (student ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Sparks, C. P. (1990). How to read a test manual. In J. Hogan & R. Hogan (Eds.), *Business and industry testing: Current practices and test reviews* (pp. 36-47). Austin, TX: Pro-Ed.
- Walsh, W. B., & Betz, N. E. (1985). *Tests and assessment* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.



附录6.1 工具和测验资源：书籍、期刊以及测验出版商

书籍和专论

- Mental Measurements Yearbooks*, updated periodically, published by the Buros Institute of Mental Measurements, University of Nebraska Press, Lincoln, NE.
- Test Critiques*, Vols. 1-7, 1984-1988, edited by D. J. Keyser & R. C. Sweetland, published by Test Corporation of America, Kansas City, MO.
- Directory of Unpublished Experimental Measures*, Vol. 1 (1974) and Vol. 2 (1978), edited by B. A. Goldman & J. C. Busch, published by Human Science Press, New York.
- Dictionary of Behavioral Assessment Techniques* (1988), edited by M. Hersen & A. S. Bellack, published by Pergamon Press, New York.
- Measures for Psychological Assessment: A Guide to 3,000 Original Sources and Their Application* (1974), edited by K. T. Chun, S. Cobb, & J. R. P. French, published by the Institute for Social Research, University of Michigan.
- Measures of Personality and Social Psychological Attitudes* (1991), edited by J. P. Robinson, P. R. Shaver, & L. S. Wrightsman, published by Academic Press, San Diego.
- Measures for Clinical Practice: A Sourcebook* (1987), by K. Corcoran & J. Fischer, published by the Free Press, New York.
- ETS Test Collection Catalogue* (1986), published by the Educational Testing Service, Princeton, NJ.
- Handbook of Research Design and Social Measurement* (1991) (5th ed.), edited by D. C. Miller, published by Sage Publications, Thousand Oaks, CA.
- A Sourcebook for Mental Health Measures* (1973), edited by A. L. Comrey, T. E. Backer, & E. M. Glaser, published by the Human Interaction Research Institute, Los Angeles.
- Test: A Comprehensive Reference for Assessments in Psychology, Education, and Business* (1986), edited by R. C. Sweetland & D. J. Keyser, published by Test Corporation of America, Kansas City, MO.

期刊

Applied Psychological Measurement
Educational and Psychological Measurement
Journal of Educational Measurement

测验出版商

- Consulting Psychologists Press, Inc., 3803 East Bayshore Road, P.O. Box 10096, Palo Alto, CA 94303 (1-800-624-1765).
- Multi-Health Systems, Inc., 908 Niagara Falls Boulevard, North Tonawanda, NY 14120-2060 (416-424-1700).
- Psychological Assessment Resources, Inc., P.O. Box 998, Odessa, FL 33556 (1-800-331-TEST).
- The Psychological Corporation, Harcourt Brace and Co., 555 Academic Court, San Antonio, TX 78204-2498 (1-800-228-0752).

附录6.2 使用申请书样本

快速识别索引 (QDI) (AKA 社会态度调查) 使用申请书

在使用QDI时,我同意如下条款:

1. QDI的版权归Fordham大学林肯中心心理和教育服务司的Joseph G. Ponterotto博士及其同事所有,地址是113 West 60th Street, New York 10023-7478 (212-636-6480)。
2. 我是一个在心理学或某个相关领域受过训练的专业人员,已经完成了多文化问题、心理测量学、研究道德规范方面的课程(或训练);我正在这样一个人监督指导工作。
3. 在使用QDI过程中,我会恪守美国心理学会和相关专业组织规定的所有道德标准,遵循由我所在的大学或专业机构制定的“人类被试研究”指导纲要。我将考虑下列(但被局限于)研究过程中的道德问题:被试同意参加实验前的知情权、记录的保密性、实验前后的恰当说明以及被试看到有关研究目的、方法、结果和含义的简要书面总结的机会。
4. 与公认的职业操守一致,我将在至少5年之内保存和保护原始数据;如果Ponterotto博士(他肩负负责监督使用、信度和效度等方面发展的道德责任)提出要求,我将把原始数据提交给他以及其他研究歧视、种族主义、男性至上主义的学生或学者。
5. 不管研究是否已经出版、展示或完成,我会把(任何涉及使用QDI的)研究结果的手稿复印件寄给 Ponterotto博士。

签名: _____ 日期: _____

姓名: _____

地址: _____

电话号码: _____

E-Mail: _____

如果是学生,主管/导师的姓名,从属关系,签名:

姓名: _____

从属关系: _____

签名: _____ 日期: _____



第 7 章

调查与问卷设计

Robert D. Goddard III

Peter Villanova

问卷和调查是社会科学研究中最常用的方法（Isaac & Michael, 1983）。这些方法的流行乃是因为时下大部分的研究，尤其是心理学方面的研究，目的都是了解个体自己的认识和自己对生存状况的信念，以及这些认识、信念和行为之间的关系。本章旨在为设计调查/问卷工具的学生提供一个基本的指导。我们按照帮助你理解调查/问卷设计与管理形成认知模式的方式来组织本章。本章从对调查法进行综述以及对调查/问卷设计过程的描述开始，这一部分将使你了解调查/问卷的方法，并为你进一步理解特定问题和选择打下基础。第二部分对调查/问卷的设计进行更为详细的探讨，旨在反映各方面的技术问题，这些问题都是你在研究中可能遇到的。最后的部分探讨问卷调查的替代方法。

调查、问卷设计与管理的基本知识

什么是调查

调查是一种从人群中收集数据进行描述或预测的方法。调查可以用来收集以全国人口为整体的信息（如10年一次的美国人口普查）或者估计一群消费者对某种软饮料的反应。调查可以采用问卷的形式、面对面访谈的形式或是电话访谈的形式。收集到的数据可以档案保存，其他研究者可以根据研究目的对数据进行分析（Zaitzow和Fields，本书第20章）。

调查目的

调查研究的目的很多，包括辅助决策、评估当前政策或措施的有效性、为某个研究项目提供信息。当然，研究人员还有其他获得信息的方法，如仅仅通过观察获得信息。但是假如所需的信息不是可以观察的行为，该怎么办呢？换句话说，研究者可以很简单地观察到行为，但对以态度为形式的信息却无能为力。我们无法观察人的态度、观点或

信念，但是我们可以通过调查去了解人的态度、观点或信念。

调查的类型

常用的调查方法是问卷法、访谈法和电话访问。问卷包含针对研究目的的条目式的调查。问卷可以单独回答，也可在调查者的安排下成组进行。调查者的工作是解释调查的目的、回答关于调查条目和程序的问题、保证调查程序正确进行。问卷的局限在于题目被预先定好了，被试不能自由地表达自己的想法。

访谈法与问卷法有很多相似之处，因为它也可以包含一套题目，以获取调查者需要的信息。但是，通过访谈，调查者可以听取被访谈者的解释，观察被访谈者的反应，这些都是问卷法难以做到的。

电话调查是通过电话进行访谈。研究者事先准备一套问题条目，根据被访谈者的回答进行追问。

职业道德问题

在匿名和保密之间进行选择是调查者经常碰到的问题。被调查者在调查中可能不希望他们的身份为人所知。满足他们这个需求可能有助于得到较为可靠的数据。被调查者作答后将问卷邮寄回研究者，这样的方式能保证匿名，因为只要信封和问卷上没有识别信息，无人知道被调查者是谁。面谈和电话访谈，因其操作特点，则无法做到匿名。

保密是保护被调查者权益时更为常用的方法。在保密的情况下，调查者能将被调查者的应答内容与被调查者本人一一对应起来，但是他不能将这个信息提供给其他人。一般来说，研究者对被调查者做出保密的承诺，将被调查者的应答材料妥善保存，仅公布集中起来的数据。研究者倾向于采取保密而不是匿名的方法，因为他们往往需要知道作答的个体是否是她所感兴趣的人。在任何情况下，无论研究者向被调查者做出怎样的保密承诺，研究者应该采用恰当的程序来兑现这个承诺。

调查的设计过程

恰当的调查设计包括几个步骤，每一步骤的成功依赖于前一步骤的做法是否有效。调查过程具体包括如下步骤：

首先，研究人员应该明确研究的目的和假设。可以从考虑是否需要收集被试的个人信息（如年龄、职业、教育）、态度、信仰、意图以及行为开始，目标是收集与研究密切相关的信息。列明细表是帮助调查者更好收集调查信息的一个方法。

明细表（Hopkins, 1978）是一个二维的矩阵表格，表格的一个维度表示研究关心的问题（如工作满意度、对监狱拥挤问题的态度和人口统计学数据）；另一个维度代表这些问题的具体表现情况。明细表格可以提供简短和准确的信息，有助于了解哪些项目合

适、哪些项目应合并或分开、如何更好地安排这些项目。

第二步，确定被调查者的特点，包括他们对调查题目的理解能力以及参与调查的可能性。抽样人群易于接近吗？如何安排才能保证样本的代表性？为保证被试人群能理解问卷的内容，问卷问题应基于何种阅读水平？McCready在第8章探讨了提高样本代表性的抽样程序问题。

第三步，设计能获得有效信息的问卷题目并决定合适的应答模式。例如，你只希望被试将选项按照喜爱程度进行排序，还是希望问卷能将喜爱程度以连续变量的形式反应出来？简单地回答“是”、“否”能满足研究要求吗？

第四步，要考虑问卷的结构，包括给被试的指导语、问卷的每个问题如何安排、哪些项目可以放在一起等。一般来看，采用从概括到具体的安排顺序比较明智。但最重要的是，问卷的安排应该以应答者为中心，让他们感到方便和友好。有时候逻辑上合理的顺序并不能让被试产生最佳的心理效果。例如，一些作者建议，在问卷的末尾要求被调查者填写人口统计学信息（如姓名、年龄、收入等），而不是在问卷的开始部分，因为提供这些信息所引发的疑虑可能会影响他们其后的应答（Frey, 1989; Miller, 1991）。

版面设计和外观与调查的施测程序有关。所以应该为较为系统的研究——包括实验室研究和调查研究——建立规程。在调查研究中，规程应该包括收集数据所需的材料、需要反映给被试的信息（如问卷的确认编码或其他数据）、向被试宣读的指导语、可以向被试提供的关于问卷项目的解释、向被试提供的关于调查程序的信息（以减少他们的困惑）。

实地调查之前，最后一步是采用预试对设计进行总体上的评估。如果预试发现还有其他问题需要考虑，就要重复整个步骤，直到调查问卷符合研究的标准。预试是一个非常有用的办法，可以帮助发现问卷潜在的问题，使问卷达到较好的质量。

不同形式的预试被调查者对问卷设计的参与程度也不同。参与型的预试要求被试作为本项目的合作人员。参加者完全知道研究的意图和不同项目的目的。当他们完成问卷，被试可能会被提示去评价问卷的清晰度、意义和偏见等问题。如果事先没有向被试说清除，则这些被试与未来的被调查者一样，以同样的操作程序回答问题，最后，将他们集中起来，进行小组讨论。小组成员包括6~8名被试和一名研究人员。以小组为单位，询问被试对调查的反应以及有关施测规程的问题。集中的小组讨论可以建立在事先准备好的访谈规程基础之上，当然，应该给被试主动提问的机会。

调查的设计问题

格式

调查应该多长？这要根据调查的目的以及调查施测的方法确定。你的研究目的决定了项目数，以保证所收集数据的可信性。

调查的形式——问卷、访谈还是电话采访——也影响着调查的长度。由于时间等限制，电话采访的时间应该在30分钟以内。面谈的时间也应有所限制，但是可以长于电话采访时间。问卷的调查最好控制在1小时之内。不受时间地点限制，要被试自选时间的问卷，应该限制在1~2小时之内完成。需要书写文字的问卷和需要被试付诸思考的问卷要比那些仅要求被试判断同意或不同意某种观点的问卷耗费更多的时间。Heberlein 和 Baugartner (1978) 对98个邮件调查问卷的研究发现，平均而言，被试需要花费半个小时来完成72个项目。

设计与布局

在调查之前，你要介绍问卷的目的，你的职业是什么，为什么要进行调查。你可以告诉被试完成问卷大概要花费的时间，说明为什么选择特定的被试来完成问卷，同时保证调查的匿名性与保密性。

题目的顺序

问卷的第一个题目就应该与调查的目的直接有关。接下来的题目顺序则因你的被试而异。一些被试可能更愿意先回答一些比较客观的问题，直到他们对调查及其目的有所了解。如果将题目按照从熟悉到不熟悉的顺序安排，被试可能感到更习惯一些。同样，对一个题目的作答会影响到对其后题目的作答，这个问题也应该考虑到。人们通常希望自己的行为保持一致，这个倾向增加了不同题目的应答之间的相关。因此，可以考虑调整调查的组份，使题目的顺序因人而异，从而使你可以在调查之后分析应答是否存在顺序效应。

标度

建立满足研究目的要求的应答标度，是研究者应该考虑的一个重要问题。区分不同类型的应答标度，以便对应答进行编码从而有助于数据分析，这是一项重要的工作。Yaffee 在第15章分析讨论了如何针对不同的数据进行数据分析，Betz在第19章描述了有关调查问卷的建构原则。

命名标度

命名标度本质上是分类性的，即所研究的变量的数值并不具有数学上的意义，而是对类别的描述。这些数值不可以加、减、乘和除，而是像标签一样对所研究的变量进行描述（Kerlinger, 1985）。命名标度经常用来收集事实信息，如人口统计学信息；对态度的测量则应有更好的标度，例如顺序和等距标度。命名标度的一个例子如下：

我的民族是（选择一个相符合的答案）：

- （1）非裔美国人；
- （2）亚裔美国人；
- （3）白种美国人；
- （4）墨裔美国人；
- （5）本土美国人（印第安人）；
- （6）其他。

顺序标度

顺序标度表示对条目的应答结果有大小顺序之差，但是不同结果之间的差异无法用数值来衡量，也不能指出不同应答之间的差异是否相同（Kerlinger, 1985）。例如：

您认为此航班的服务（选择一项）：

- （1）比我上次乘的航班服务好；
- （2）与上次的航班服务同样好；
- （3）没有上次的航班服务好。

等距标度

调查中最常用的等距标度是Likert或者累积评定尺度。Likert尺度经常用于态度或观点调查，要求被试他们对题目的同意或不同意程度进行估计。应答经常包括5个或者更多的分类。例如：

这个公司的工作条件今年比去年更好。

- （1）完全不同意；
- （2）不同意；
- （3）不能确定；
- （4）同意；
- （5）完全同意。

回答者选择一个且只是一个答案。这个调查具有绝对性和排他性，意味着回答者不



能选择多于一个的类别，且可能的选择包含了所有可能性。

有关Likert是顺序还是等距标度存在一些争论，相当多的研究已经证实它最接近等距标度（Spector, 1992）。

题目的特点和回答的形式

有几个特征可以区分题目的好坏。主要的区分取决于需要得到的信息种类和应答者样本。题目某些特别特征会对应答错误和偏见产生较大的影响（Bradburn, 1982）。

定义术语

访谈者应该给任何可能模棱两可或含糊的概念或习语提供定义。这可以通过添加附带的描述来完成，例如，“我支持死刑（也就是说，我认为死刑是对那些犯下某种重大罪行的人的恰当惩罚）。”不要假定目标样本的所有成员对一个概念都有一致的理解。应该给任何有问题的词和习语提供定义。

问卷面临的一个挑战是如何让题目对不同应答者具有相同的意义。例如，在访谈者事先没有定义术语“适应”的情况下，题目“在何种程度上你相信下列雇员适应他或她正在进行的工作”，对“适应”的理解因人而异。一个应答者可能认为它的意思是某人工作环境中与他人的融洽与否，另一个应答者可能理解为某人是否具备某项工作所必需的教育程度和技能，第三个应答者则可能认为这个题目问的是某人是否满足了工作要求。Morrow、Eastman和McElroy（1991）发现，由于对“工作态度”这个概念不熟悉，应答者对5个不同的工作态度量表的应答产生很大的偏差。

在不同的应答选项中使用词语标度可以使得题目更为清晰。Bass、Cascio和O'Connor（1974）以及Spector（1976）针对词语标度（如偶尔、很少、经常、几乎不）在应答者感知的强度方面的差异，进行了研究，提供了有用的信息。当回答包含应答者对行为的评价以及这些评价需要以等距数据表示时，词语标度就变得尤其重要。

开放与封闭的问题

一般来说，问卷的问题越具体，不同被试对问题的理解就越有相似性，这样对问题的回答就会越准确越具有可比性。相反，开放的问题允许被试给出关于每个具体题目的不同解释。不过，如果要限制被试的应答方式，可以在问卷中指出“关于下面的主题最多罗列出3个事项”。封闭性问题包括一个陈述性问题（例如，“哪一个问题对你来说最重要？”或者“将这些问题对你来说的重要性按从最低（1）到最高（4）排序”）和跟随其后的几个选项，例如：

____ 预算赤字

____ 经济增长

——医疗保健

——社会公正

开放的问题经常在调查研究的开始阶段使用。当你对于一个现象尚有不能透彻的理解，可以采用开放问题，通过应答者的回答，尽量覆盖现象的方方面面。当对现象有较好的理解后，随后的研究可以采用封闭的问题形式。

开放的问题也可以用于追问应答者对某些封闭式问题的回答，从而提高被试对调查题目的理解。例如，在1983年的底特律地区福利态度研究（Detroit Area Study）中，大多数的回答者同意“政府正在做很多本应留给个人或私人企业来做的事情”这样的观点。然而，当随后问到“你认为什么样的事物应该留给个人或者私人的企业来做”时，则有大约25%的回答者无法回答这个问题或者不能举出实例（Converse & Presser, 1986）。

一般来说，封闭的问题需要更少的施测技巧，有利于数据的分析，同时也可以增加应答间的可比性。不过，使用封闭的问题要假定应答选择涵盖了人们观点的范围。它涉及是否对应答选择做出限制。在一个从绝对自由到严格限制的连续范围内，在“自由”端是开放问题，诸如“列出那些你相信政府能做好的事”；在另一端，意味着对回答严格限制，像律师经常希望证人提供的那种“是”或“否”的回答，此时的问题则是如“你觉得政府有益吗？”在这些极端问题的中间部分是容许应答者自由回答的问题，如，“列出不多于3件事……”“将下列条目按照顺序……”以及“从这个清单中选择……”。

问题的筛选

不能假设所有的应答者都理解你的问题，并都形成关于问题的意见。你如果要求说出意见而没有提供“没有意见”的选项，就可能会降低调查的准确程度。除了提供“没有意见”的选项，你还可以通过提出一个筛选问题来询问被试对某问题是否有自己的看法，若有，指出其看法：

不是所有人必须对该问题提出自己的意见，如果没有意见，请回答“没有”。问题为“联合国所作所为可以缓解波西尼亚的压力与痛苦”。你是否有关于此问题的观点，如果有，你是否同意或者不同意上面的陈述？

一些研究者不喜欢“不能确定”或者“没有意见”这种选择，而是迫使被试或多或少地提出自己的观点。另一些研究者利用所谓“中立”回答来决定问卷题目的重要性。如果许多的回答者表明他们的态度是“中立”，则表明这个问题对他们来说不重要，或者他们没有足够的信息来对自己的立场作判断（Orlich, 1978）。例如，在前面提到的例子中，如果是刚到公司一年之内的职员则会回答“不能确定”或“没有意见”。不过，“没有意见”选择应该慎用，因为研究者发现，10%~30%的回答者以没有意见作为回答

(Schuman & Presser, 1981)。

同样，如果你的目的是测量信仰或者态度，你可以考虑忽略掉中性的应答选择。例如，你可以问：“你认为堕胎服务在这个国家应该更难获得，更易获得，还是保持现状？”这样一来，许多对这个问题没有什么看法的人便可以选择保持现状作为选择。不过，如果你将选择限制在“更易”和“更难”两项，接着提出“你在多大程度上”这样认为，则可以将在这个问题上持有温和观点的应答者们区别出来。

另一种情况是，迫使被试在两个似乎都可选择但实际上相互矛盾的选项中进行选择。例如，题目可能要求被试同意或不同意如下的观点：(a) 政府应该保证每个人的医疗保健费用。(b) 医疗保健是个人应当承担的开销。被试可能对a和b都表示同意，这就不能区分被试对两个选择的偏好。一个好的解决办法是将两个问题合并成一个问题：(c) 你认为是政府还是个人应为医疗费用负责？

单一对象，单一含意

避免让被试同时评估多个对象：例如，问题“你是否同意如下的陈述：应该让教授和学生参与那些影响他们工作的政策的制定？”，除了需要对“工作”和“政策”作更为精确的定义，“教授”和“学生”这两个对象的同时存在也会令读者不知所措。一个较好的解决办法是将它们分成两个题目来问。

题目应该含意单一，易于被试理解。例如，问被试是否同意/反对“不应该要求教授课前点名”的说法会导致混淆。一个反对的应答可能意味着“我不同意教授不应该被要求课前点名。”最好将这种题目按正向语句的方式提出：“教授应该被要求进行课前点名。”

默许或人为的顺从

如果调查的目的是描述别人的观点和信仰，则问题不应该具有诱导性。应该避免诱导性的问题。例如：“你是否支持克林顿总统的先进的健康计划方案？”问题中假设了此动议是“先进的”，而且回答者会迷惑，到底是支持克林顿还是支持克林顿的健康计划。另外，对这个问题的提出还有赖于对“你对克林顿的健康计划是否熟悉”、“你对此计划是否有自己的建议”等这样问题的回答。同样，为了避免在问题的开始部分就具有诱导性，应该避免“难道你不同意……吗”这种会显著提高同意倾向的提法。

问卷调查的替代方法

访谈

访谈大概是在时间上和经济上花费最大的调查方法，但访谈有很多优点：（a）访谈可以获得用其他方法不能估测到的被试态度；（b）通过讨论能够使被试对问题有深刻理解；（c）鼓励自由表达；（d）观察和记录非言语的交流；（e）发现个人情况、态度、信仰和感觉等用纸笔调查可能忽略的信息；（f）保证较高的被试参与率；（g）使主试能够追问或者按调查条目的顺序进行；（h）有助于书写阅读困难或者有视力缺陷的被试完成调查（Orlich, 1978, p.8）。

另一方面，访谈可能需要经过广泛训练的访谈者。这有几个原因。其一，访谈者可以通过影响受访者对调查的态度而直接影响访谈结果；也可能因其展示调查内容的方法不当而失去潜在的受访者。其二，访谈者须能够指导受访者如何回答采访问题。最后，访谈者必须以标准的方式进行访谈，以避免应答的偏差（Fowler, 1993）。

访谈者需要得到受访者的合作。出色的访谈者以干练、自信和果断博得受访者的合作。他们也能够主动调整自己的介绍过程，使受访者得到足够的重视。

培训可以帮助访谈者，使他能够以规范的方式来展示调查记录，并在被试反应不完全时提示他们，进行追问。培训也能够帮助访谈者学会正确、连贯地记录应答结果。尤其重要的是，培训能帮助访谈者处理那些需要对应答进行理解和归类的开放式问题。

电话访谈

你或许在晚间新闻里听到过这样的报道：“53%的被调查者认为，美国的总统工作做得很好。”在如今的高科技社会，有个从庞大的人口收集信息的方法，这个方法就在我的手边上。现在，我们可以“接通”他人（如果他们电话），通过与电视观众分享信息的方式来影响国策。

电话调查通过使用精心设计的问题，以获得有关于人们的态度和行为的种种事实。

正如Fowler（1993）提到的，电话采访，与其他方法相比，有很多的长处：（a）潜在的低费用；（b）对大范围的人群可以采用随机—数字拨号抽样；（c）可以采访到某一特殊人群；（d）收集数据可以在较短时间内完成；（e）减轻寻找、管理访谈者的负担；（g）在特定样本下有较高的应答率（p.65）。

如果被试集中在某一个地方而且调查不需要较多的培训，电话采访可以比其他的方法需要更少的资金和人员。对样本的选取可以通过随机—数字拨号来进行。这个方法保证了样本的代表性。当然，没有电话的家庭会被排除在外，不过这种情况比较少见，研究者对此并不是很担心。但当面对某些研究问题，研究者必须正视电话调查的有效性。

如果你希望调查无家可归者对民生问题（如住房和就业）的看法，你可能无法得到一份名单或者从电话号码簿中抽样。此时，你就得考虑别的数据收集策略。事实上，用电话调查去研究一个社区对周边发展的态度，无疑是非常合适的。在这种情况下你就需要得到一个电话号码簿，并从中抽取受访者。关于抽样的策略，可以参考本书其他章节（见 McCready 所著第8章）；关于电话调查，参见本书第11章（Chen）。

总 结

本章为调查和问卷提供了大量的信息，以供您参考。但是由于篇幅有限，一些问题不能充分讨论。我们建议你考虑其他的资料来源（如下面提到的参考文献），以帮助您提高设计调查的能力。我们希望这一章为您理解其他相关材料打下了基础。

推 荐 阅 读

对本章阐述的问题有进一步兴趣的读者可以参考如下的书籍：Fink 和 Kosecoff 所著《如何进行调查》（*How to conduct Surveys*, 1985）；Guenzel, Berkman 和 Cannell 所著《常用访谈技术》（*General Interviewing Techniques*, 1983）；Payne 著《提问的艺术》（*The Art of Asking Questions*, 1951）；Rosenfeld, Edwards, 和 Thomas 著《改善组织化调查：新方向、方法和应用》（*Improving Organizational Surveys: New directions, Methods, and Applications*, 1993）；Sudman 和 Bradburn 著《提问》（*Asking Questions*, 1982）。一些论文也提供了有用的信息：如，Gibson 和 Hawkins（1968）；Hunt, Sparkman 和 Wilcox（1982）；Wiseman 和 Billington（1984）以及 Yu 和 Cooper（1983）。

参 考 文 献

- Bass, B. M., Cascio, W. F., & O'Connor, E. J. (1974). Magnitude estimations of expressions of frequency and amount. *Journal of Applied Psychology*, 59, 313-320.
- Bradburn, N. M. (1982). Question-wording effects in surveys. In R. Hogarth (Ed.), *New directions for methodology of the social and behavioral sciences: Question framing and response consistency* (pp. 65-76). San Francisco: Jossey-Bass.
- Converse, J. M., & Presser, S. (1986). *Survey questions: Handcrafting the standardized questionnaire*. Newbury Park, CA: Sage.
- Fink, A., & Kosecoff, J. (1985). *How to conduct surveys*. Newbury Park, CA: Sage.
- Fowler, F. J. (1993). *Survey research methods* (2nd ed.). Newbury Park, CA: Sage.
- Frey, J. H. (1989). *Survey research by telephone* (2nd ed.). Newbury Park, CA: Sage.
- Gibson, F., & Hawkins, B. (1968). Interviews versus questionnaires. *American Behavioral Scientist*, 12, NS-9-NS-11.
- Guenzel, P., Berkman, R., & Cannell, C. (1983). *General interviewing techniques*. Ann Arbor, MI: Institute for Social Research.

- Heberlein, T. A., & Baumgartner, R. (1978). Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature. *American Sociological Review*, 43, 447-462.
- Hopkins, C. D., & Antes, R. L. (1978). *Classroom measurement and evaluation*. Itasca, IL: Peacock.
- Hunt, S. D., Sparkman, R. D., Jr., & Wilcox, J. B. (1982). The pretest in survey research: Issues and preliminary findings. *Journal of Marketing Research*, 19, 269-273.
- Isaac, S., & Michael, W. B. (1983). *Handbook in research and evaluation* (2nd ed.). San Diego: EDITS.
- Kerlinger, F. N. (1985). *Foundations of behavioral research* (3rd ed.). New York: Holt, Rinehart & Winston.
- Miller, D. C. (1991). *Handbook of research design and social measurement* (5th ed.). Newbury Park, CA: Sage.
- Morrow, P. C., Eastman, K., & McElroy, J. C. (1991). Concept redundancy and rater naivete in organizational research. *Journal of Applied Psychology*, 21, 219-232.
- Orlich, D. (1978). *Designing sensible surveys*. Pleasantville, NY: Redgrave.
- Payne, S. (1951). *The art of asking questions*. Princeton, NJ: Princeton University Press.
- Rosenfeld, P., Edwards, J. E., & Thomas, M. D. (1993). *Improving organizational surveys: New directions, methods, and applications*. Newbury Park, CA: Sage.
- Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys*. New York: Academic Press.
- Spector, P. E. (1992). *Summated rating scale construction*. Newbury Park, CA: Sage.
- Sudman, S., & Bradburn, N. M. (1982). *Asking questions: A practical guide to questionnaire design*. San Francisco: Jossey-Bass.
- Wiseman, F., & Billington, M. (1984). Comment on a standard definition of response rates. *Journal of Marketing Research*, 21, 336-338.
- Yu, J., & Cooper, H. (1983). A quantitative review of research design effects on response rates to questionnaires. *Journal of Marketing Research*, 20, 36-44.



第 8 章

取 样 过 程

William C. McCready

取样是一种解决特定问题的方法，因为我们无法从简单一瞥中获知全部事实，而是一次只看到一个事件。取样的目的是提供系统的规则和方法，来检验样本能够在多大程度上代表研究的事实。

一方面，让我们试想第一批研究人体内部构造的人。他们不知道将发生什么事，也没有可用的解剖图。早期作解剖研究的学生研究尸体，详细记录下人体的循环系统和器官，这些早期的科学家们最终提供了一幅人体图——它对西方医学的发展起了重要作用。利用这张图，外科手术先驱很快发现，所有的人在解剖上都十分相似。外科医生能够将这张通过研究人群中很小样本所得出的图推广到整个人类，这是因为就解剖结构而言，诸如循环系统和器官分布，个体间的差异是极小的。

另一方面，想想如果问某人10个关于生活目标的问题。世界上其他人的回答也会和他一样吗？为什么回答某类问题只需要很小的样本，而回答另一些问题却需要很大的样本？你能否不通过常识而说出何时需用何种方法？为何一些科学家采用取样方法而另一些科学家不采用它？有一点需要记住的是，取样是多种研究工具中的一种，并且只能用于特定目的。这种工具就是本章要讨论的主题。

取样和科学方法

代表性

取样的最终目的在于获得一小组“事件”，它们能准确地代表更大组同样“事件”的属性和特征。这个小组称为样本，大组的称为总体。如何选择样本，其关键是看样本对总体具有多大的代表性。无论是使用控制组和实验组，还是作社会调查，你都会希望选取的样本能代表研究的总体。取样是测量小组对大组的代表性的方法。

十分重要的一点是，我们无法得到具有完全代表性的样本。样本永远不可能是总体

的完全复制。科学中的代表性是一种测度和近似值，而不是复本。我们应选择一个在一定条件下能最好估计群体的样本。取样的规则和方法针对的是如何选取样本，使其是对总体的可靠和精确的估计。举个例子，假设我们想知道有多少人会对两种情境下的灾难受害者提供帮助：第一种，受害者没有得到灾难的预警；第二种，受害者事先得到预警但没有采取有效的措施。我们可以设计一个实验，检验它们的差异，但是这个实验不会是事实的完全重复。取样的规则和方法能够提供一种指导，以便样本能够更大程度地接近事实。

可重复性

科学一个最为重要的特点是，你制定的调查研究能够被另一个科学家重复，并得到与你相同的结果——这称之为可重复性。不同的学科有不同的研究方法，但不管你是一个在实验室中研究基因的生物学家，还是一个研究家庭生活和社会暴力之间关系的社会心理学家，你都希望自己的研究发现具有可重复性。

社会科学家，例如心理学家和社会学家，处理的研究问题与早期解剖学家遇到的问题相比，包含了更多的变异。人类行为具有难以置信的复杂性并且变化多端。如果你觉得可以简单地通过选用大样本来获得所有变量，那你就错了。

实验方法

最有效的研究工具之一是实验。许多类型的实验需要大样本，但有些实验并不需要。事实上，有时小样本实验更加有效。美国心理学家B.F.Skinner（1966）说过：“上千小时研究1只老鼠胜于每小时1只地研究上千只老鼠，或是每10小时1只地研究上百只老鼠。” Sigmund Freud从心理学的另一端，也深入研究一些被试，而不是肤浅地研究大量被试。当验证一些重要原理时，实验特别有用；当有理由相信，对所研究的问题而言，所有被试都十分相似——例如研究人类的循环系统，对一个被试的细致观察尤其有用。

实验研究经常采取将被试随机分配到实验组和控制组的办法，对控制组的成员不施加实验操纵。两组被试通常在一些变量上匹配，如性别、年龄、体格等，这样比较两组的结果，可以反映出施加于实验组身上的操纵所产生的影响。

总体估计

当然，也存在一些无法用实验研究的问题。例如，你没有要检验的理论，或是你怀疑抽取的被试之间存在太多的差异。有时研究的问题包括探知更大群体中的条件，例如，“社会中的人如何看待环境问题？”对这种问题，需要以社会群体作为样本。

本章概况

取样为不同的工作提供了不同的研究工具，取样的方式取决于要做的工作。有实验的取样和社会调查的取样；有选取合理数量被试的取样和反映总体估计的取样。在下一部分中，我们会讨论取样的类型和工具，并就何时用何种取样方法提供一些指导。

有两大类取样，非概率性取样和概率性取样，各自有各自的用途。我们将首先讨论非概率性取样，因为心理学家经常使用这种方法为实验挑选被试。

非概率性取样技术

关系

C.James Goodwin (1995) 写道：“当然，我们希望这些研究结果不只对参与实验的被试有效。但研究者假定，无论被试是如何被挑选的，如果研究的相关关系很高，那么这结果对其总体的大部分人也适用。”这个论断很好地描述了变量之间的关系效力。关系很强，如服从于权威合法性之间的关系，则不管试验被试是如何选择的，这种关系都可推广到总体。在这种研究中，许多类型的非概率性取样十分有效。这些取样有多种名称，包括判断型、事后型、方便型、有目的型和匹配取样。

被试库

有人会说，非概率性取样不是真正的取样，它包含了被试选择的系统效应。举例来说，大学心理系一年级的学生通常会参加心理学实验，这对他们来说是普通的实习。这些学生便形成了“被试库”。虽然学生被试可以从库中随机抽取，但库本身并不一定会代表大的总体。然而，就实验目的而言，挑选出的被试确实能代表被试库中所有适合参加实验的被试群体。

选取和分配被试

为保证实验伦理，被试应该是真的自愿者，不带任何强迫性。被试不允许挑选实验和条件（例如，选择到实验组还是控制组），而是随机分配。这些随机化程序防止因被试以前的经历或对实验的认识以及其他潜在原因带来偏差。偏差是取样错误的原因之一，因为先前知识或分配产生的偏差，在选取过程中是可以避免的。研究者通常利用他（或她）对样本的判断来设置选取过程。举例来说，如果你研究的是慢性疲劳症，你就不会对在访谈中表示从未感觉到疲劳的人产生兴趣，你的专业判断会在样本中排除这类人。

匹配样本

理想匹配的样本由同卵双胞胎组成，在许多国家，双胞胎都是实验者关注的目标。将双胞胎中的一个放入控制组，另一个放入实验组，往往消除了所有无关变量，只留下我们研究的实验变量。

在大多数实验中，在显著的特征上对实验组和控制组的被试进行匹配，来模仿理想的状态。匹配通常在选取样本时就完成了。开始时记录被试特征，此时就开始了匹配。除了实验变量，实验组和被试组越相似越好。当我们发现两组结果有差异时，我们就可以认为，这个差异是由实验造成的，而不是由组间的无关差异造成的。

非概率样本中的取样“误差”：一个命名错误

在社会科学研究中，“取样错误”是一个常常被误解的概念。更合适的说法应该是，“样本差异”。样本差异是不可避免的，即使多次重复取样过程，差异还是会存在。称其为“错误”是不合适的，因为样本差异是取样总误差中最小的成分。另外两个成分是样本偏差和反应效应，这些通常都会很大。样本偏差来自于取样的计划有误，反应效应则是被试报告的行为和态度与真实的行为态度之间的差异。

此处的问题是，你能否从重复选取的样本身上得到同样的测量和统计值。对这个问题我们在下面讨论概率取样时还会提到。我们通过累计重复测量产生的变异，并取其平方根来测量样本差异。换句话说，平常所认为的样本“错误”其实是重复取样的标准差。

我们能计算这个标准差，但很难作出解释，因为被试被挑选出来的几率是未知的。我们将在下一部分做更多的探讨。但因为很难定义非概率性样本的“总体”，所以我们很难对选取过程作概率性的说明。为此，对从被试库中征募被试的取样，不常讨论样本误差。

样本大小

样本大小是理解非概率性取样和概率取样的桥梁。事实上样本大小这个概念如同“误差”一样经常被误解。实验需要多大的样本？在实验研究中，与实验的试次、实验条件数以及你想做的操纵数有关，而与估计更大群体的值无关。但是，当对一项诸如社区研究计划作概率性取样时，情形就倒过来了，因为此时对总体参数的估计精度就变得极为重要了。

对非概率性取样来说， N （样本大小）应该足够大，可对实验变量的变化进行可靠的测量。对初学者来说，一个好的办法是在阅读研究文献时记录下研究所选取的样本数。换句话说，当你回顾本领域的文献时，在设计自己的实验之前，注意别人报告的样

本数，它们的平均数可以作为你的参考。最近3年的实验报告，可以作为粗略的指导。典型的实验样本范围在50~200之间。关于样本容量的其他细节请见第15章。

很明显，操作被试的费用所需也是选择样本大小时应该考虑的方面。你有多少助手？被试做完实验要多久？实验一次可以做多少被试？你是否为实验准备了大量的资金？这些都是需要考虑的重要问题。通常可以通过询问做过此类实验的同事得到相关的信息。如果无法得到信息，仔细阅读文献，或许可得到与某些研究者进行联络的信息。

概率性取样

另一个取样程序是概率性取样。在这种取样中，一个样本的每个成员都有一个已知的被选取的概率。这些概率可能并不相等，但却是已知的。

样本、总体、分析单元

样本是通过一定方法从总体中选取出来的一部分，研究者希望通过小组样本来代表总体。总体可以由任何元素组成，可以是人、家庭、学校、城市、罪犯、毕业生、商业交易、事件、决定、任务或任何类似的元素。设计问卷的一个重要任务是定义研究的总体，在此过程中重要的一点是为研究任务确定分析单元。

让我们通过一个典型例子来看看确定分析单元的问题，调查一下一个社区在过去几年中有多少人使用了医疗设施。有关部门会记录就诊次数，但除非每次记录都附有个人信息，否则我们不知道究竟有多少人。既然如此，研究者需要决定分析“已通过医疗设施做了治疗的事实”还是“受到治疗的个体”。第一种分析单元（治疗事件本身）已经有相应的数据，第二种分析单元要收集新的数据（接受治疗的人）。事实上，完成这项调查需要新的数据，为每个被试建立独一无二的个人标识（使用过或没有使用过）。

概率理论

概率理论对取样的重要贡献是它确保了研究者知道样本中个体被抽取（样本中某个事件发生）的几率、机会或概率。这些信息为什么有用？因为如果你知道了某个个体被抽取的几率或概率，你就能推知它未被抽取的概率，并对相同实验设计的重复测量样本之间的差异有更精确的测量。通过这种方法，你能为代表这一概念赋予数值。例如，假设我们抽取了一个社区的样本，发现样本中25%的被试同意建一个新学校，75%的人不同意，我们想确认这个样本是否准确代表了社区总体的态度。如果我们选取另一个样本怎么样呢？多大程度上会得到相同的结果？样本变异，即样本之间差异程度，是对这个问题的回答。你可以通过概率计算出数值。对这些数值，你可以进行比较、分析和重复。但你只能对概率性取样进行这种计算，这也是为什么有些研究（如那些需要总体估计的

研究)不采取非概率性取样的原因之一。

初学者应该记住,几乎所有社会科学统计都是针对两个目的之一的,即我们试图测量事件之间有多接近或有多疏远。前者称为“集中性”,后者称为“差异性”。另一种表述方式是大多数社会科学通过一系列我们称之为变量的东西,来研究两个主题的相似性或差异。相似性或差异是否重要,通常由我们的理论表述来决定。取样设计的任务是提供一个框架或流程,来保证另一个研究者能重复我们挑选被试的方式,这样我们的发现就不只适用于一个样本了。

概率理论是数学、统计学、哲学中一个特殊的领域,它十分复杂。概率用于测量所有类型事件发生的可能性。新闻媒体为许多体育运动提供竞买点数是为了刺激赌博,而不是为了传达信息。大型集团的市场销售活动是基于对一组具有某种特征的人进行某种活动的概率测量。你的保险率也是通过概率方法计算的。概率经常决定了你的纳税申报单是否需要得到审计。日常生活中概率时刻围绕在我们身边。我们倾向于接受这样的表述,“比起飞机失事来,你更可能被雷电击中而致命”,并不知道这些测量从何得来的。

对概率理论和取样,重要的是要记住,这些理论为要作反应事件或选择被试提供了一种方法。一个选择有一个值,其好处是可以对这个值进行操纵或与其他值进行比较。如果没有这些特定的值,我们不可能准确知道,用同样的取样方法重复实验程序,会产生什么结果。没有这些知识,就无法精确地进行重复;不进行重复,就不是科学的方法。

概率取样最常见的用途在于调查研究中,因为调查通常以使用样本来测量总体的属性、性质和参数为目的。在某种意义上,尽管实验的小样本可以通过非概率方法获得而不牺牲准确性,大样本最好是通过概率方法选取,这样我们就可以测量它们的变异了。对这种现象,Kalton(1983)有较好的表述。他观察到,如果将概率性的小样本和非概率性样本做比较,概率性样本的差异会大到使样本失效;增加概率性样本的大小可以减少样本变异和使这些样本变得有用。

另一方面,因为非概率样本的样本变异无法计算,由取样失误所引起的偏差会与样本量同步增长,所以非概率取样适用于小样本的情况。在某种意义上,这就好像航海的不精确性和已经航行的路程一样。在短距离的时候,不精确性会导致几百码的小误差。在长距离时,同样的不精确性会导致更大的错误,你将偏离目的地几英里而不是好几码。用概率方法选取大样本的被试,可以保证我们得到这样一个结果:随着样本数量的增多样本变异会减少。

取样框

总体中每个成员的被抽取的概率清单称为取样框。取样框不仅仅是总体中的所有成

员；它是所有成员被作为样本选中的概率清单或序列。在简单随机取样中，总体的每个成员都有同样的被选几率，但计算不同的成员的概率可以有不同的方式，这也就是取样框的含义。

简单随机取样

简单随机取样是指总体中成员都有相同的被选概率这种抽取方式一步到位。例如，如果我们想要从总人数为7 500的社区中抽取一个样本，我们会使每个人具有 $1/7500$ 的抽取概率，事实是这样吗？如果在样本中包含了所有的婴儿与儿童，答案是正确的，但我们也许不希望这么做。如果有人口普查数据，就可以知道每个年龄组的人数，我们将用成年组（不管如何定义成年人）的人数代替7 500，这样使每个成年人的被抽样的概率为 $1/$ “成年人总数”。但是，我们通常没有人口普查的信息，因为社区人数太小了或者人口普查的数据太老了，那么就要用别的方法来得知每个成年人的被抽样概率。可能需要取样看年龄分布，然后再挑选出研究所需的成年人。这里关键的一点是，当我们面对现实研究时，即使是简单随机取样也会立刻就变得复杂起来。

系统化取样列表

简单随机取样的一个变式就是从成员列表中随机取样。大多数情况下，这是真正的随机取样，除非列表中有太怪异的周期、聚类、分组或分裂。从列表中随机取样，首先将列表成员的总数除以你准备研究的样本大小，来计算取样份数（分成 n 部分）；接着随机选择一个开始点，从此开始每隔 $n-1$ 个选取一个元素。这样你就能创建一个简单随机取样元素列表。例如，假设你有一份300名学生的班级花名册，现在你想要选一个30人的样本做实验。将300人分成30组，每组为10人，建立样本分数。随机选择一个开始点，然后每隔9个挑选一个学生（如果想要25个学生的样本，你要从每组12个学生中挑选）。随机开始点可以通过将学生从001 ~ 300编号，并使用随机数字表或随机数字产生程序获得，这种程序在许多科学家的计算机里都有，挑选呈现在001 ~ 300间的第一个数字。如果学生列表具有你的计算机可用的格式，在随机选择之前将姓名按字母排序混在一起也是一个好方法。有时列表成员会有自己的标识号，能被随机数字产生器读出，并且能被用来选择开始点。

区组配额取样

严格来说，配额取样是一种判断取样的非概率方法。它被普遍使用于早期调查研究中。访谈者会被告知以某些特征选择一定数量的被试。例如，要求挑选30名被试，这些人中要有15名女性，15名男性，其中7名男性应该在40岁以下，8名在40岁以上，8名女性应在40岁以下，7名在40岁以上。

随着概率取样的使用, 研究者们将概率选择的一些方面同以前的配额选择方法结合起来, 形成一种折中方法。先用概率方法选出区组。一旦选择好区组, 实验者会用配额方法从组块中选取个体被试。一些研究者在小成本的个人访谈中仍然偶尔使用这种方法。

全概率取样

个人访谈中区组配额取样和全概率取样通常的差别在于, 全概率取样在区组选好后对个体被试的选择仍采用概率性选取。如一个访谈者一旦到达一家住宅, 他或她会问应门的人家中有多少人, 然后用统计的方法随机从这个清单中抽取被试运用Kish表的一些变式来决定哪个人是调查对象。以取样的先驱, Michigan大学的Leslie Kish命名的Kish表, 只是一份随机产生的数字表, 它告诉实验者, “如果家中有两个人, 选年龄大的那个”, “如果有三个人, 选最年轻的那个”, “如果有四个人, 选第二大的那个”, 诸如此类。这种方法成功的秘诀在于这些表是随机产生的, 对访谈者访问的每个家庭来说都是不同的。例如, 下次实验者看表, 发现跟上面所说的不同, 我会看到“如果家里有两个人, 选年轻的那个”, “如果家里有三个人, 选中间的那个”, “如果有四个人, 选第二年轻的那个”, 诸如此类。通过这种方式, 我们保证家庭中每个成员有同等被选取的概率, 从而实现了全概率取样方法的目标。总体中每个个体有相同被选取概率的样本通常被成为epsem样本, 意为“每个成员有相等被选择概率”。

另一种在选定的家庭中选取被试的方法是找“最近过完生日的被试”。在群体中生日是随机分布的, 这种方法提供了一种合理的随机选择家庭被试的方法。

随机拨号取样

现在的调查很多都是用电话来做的, 而且越来越多的调查是基于计算机程序(计算机辅助电话访谈, 缩写为CATI)。在美国等一些国家中, 电话平均覆盖率达95%, 电话调查通常用随机拨号(RDD)的方式完成。典型的做法是, 首先估计一个地区代码有多少家庭电话号码, 然后按比例随机选取一部分。

今天, 随着电信技术的快速发展, 例如移动电话、呼机、传真、新的地区代码之类的, 研究者越来越难跟上这些技术。所以许多使用CATI的研究者从诸如调查取样公司或Genesys公司等卖主手中购买号码。这些公司根据你的要求, 提供所有的取样所需的统计工作。对此, Chen在第11章中有更详细的描述。

样本大小

关于样本, 学生通常会问的第一个问题是: “你如何知道需要多大样本?” 对这个问题的回答可以很复杂, 需要了解我们已经提到的一些要点。但从直观上看挺简单, 因

为看起来样本大小和样本品质间存在正相关——难道不是越大越好吗？这个问题不一定！记住这一点绝对是至关重要的。在取样中一个突出的问题不是“样本有多大”，而是“样本中的元素是如何被选择的”。

虽然看起来没有直观意义，但要用同样数目的被试可以保证相同的测量准确性和精度，无论这个总体是整个国家、特定的州、城市，还是一个更小的社区。换句话说，如果一个调查，在美国成年人总体中选取了1 500个被试，来保证精度保持在3%范围内，置信度为95%，在以克利夫兰市为总体时，也选取1 500个被试，我们同样可以达到相同的统计精度。

“大数定律”（Mosteller, Rourke, & Thomas, 1961）描述了其中的原因。这个法则指出，随着样本数量的增加，任何测量比例将迅速接近估计所代表的真实比例。例如，增加样本容量从500~1 000个，将精确度范围从7%提高至4%，并且增加容量到1 500会使精度范围提高到3%。但在此之后，缩少的范围将十分小，不值得增加额外样本。例如，将样本翻倍至3 000个被试，精度范围只是到了1.5%。

如想以更直观的方式来考察大数定律，请想像一下下面的实验。为了庆祝美国国庆节，清空了本地公园的游泳池，填上乒乓球；一些是红的，一些是白的，剩下的是蓝色的。我将奖励能最好估算出红、白、蓝球比例的人，给他们500美元。如果你选取50个球，并看到11个（22%）是红的，19个（38%）是白的，20个（40%）是蓝的，你会很有信心得奖吗？不会的。因为如果另一个人做同样的实验，他或她可能得出另外一个不同的比例来。

然而，如果你选取了500个球，你很可能会更有信心得奖，如果你选取了1 000或1 500个球，你会更肯定了。然而，如果你花费了更多的时间去选取3 000或6 000个球，你将发现，结果与选取1 000或1 500个球并没有多大的区别。你会觉得这并不值得那样努力，去选到3 000或6 000个球。在选取1 000或1 500个球后，你应该可以对估算的准确性很自信了：这就是大数定律在起作用。

聚类取样

一些总体有一些相似的组成成分，例如学生，你可以提取类别样，接着在类别中选取单个样本。例如，想要抽取学生，我们会首先随机选取学校，因为在那里很容易找到学生；接着我们在随机选取的学校中抽取学生，并清楚知道样本中学生的被选概率。这样做的优点在于，比起随机抽样家庭来选取学生，这种方法所需的花费和精力相当少。

分层取样

根据非地理性的特征可把总体分为许多分组。例如，我们能决定将总体分为4个组；40岁以下的男性，40岁及其以上的男性，40岁以下的女性，40岁及其以上的女性。我们

之所以决定这样做，是因为我们对40岁上下女性的生活非常感兴趣，想集中取样，以获得更多的有关信息。

分层取样并不总是合适的，要慎用。Sudman (1976) 对分层取样可能的弊端有一个精彩的描述。通常有三个原因使你选择分层：第一，层次本身是你感兴趣的问题，这可能是真的，尤其是在人口统计学意义十分重要的调查工作中；第二，你可能有关于层次之间变异或特征的先验信息；第三，分层能节省经费，但必须谨慎，这种由于费用原因而采用的分层不一定能产生费用少并且准确的样本。

多层取样

大多数国家级的调查样本是多级样本，无论数据由个人访谈还是CATI访谈收集。这意味着取样分几步完成，从自然聚类开始，到个体被试的选取结束。概率被分配到选取的每个阶段的每个单元，并能联系起来为最后的样本中每个成员的选取提供可知的独一无二的概率。自然地地域聚类，如人口普查小册子和地区代码，都被用来为总体分类，在每个类别中再按取样比例用来挑选个体。

例如，如果你居住在一个小镇上，这个小镇只有很少的机会会被抽取为样本（许多好的国家型调查样本大约有1 500名被试，一般去掉了14~16个州，因为这些州里的小镇无一参与了最后的选取）。然而，如果你们小镇在早期阶段参与了选取，比起一些住在大镇子里的人，你将有更多的机会被选取。

许多初级研究者不需要知道多级取样。然而，如果你有机会参加一个用这种技术的项目，你将学习到取样的实际技术。学习取样技术的最佳途径是给一个专家老手当学徒。这是一种改进、精炼和代代相传的技能，也是学习该技术的最好方法。

其他特殊的取样设计包括纵向研究取样、群体取样、累积取样、聚合交叉取样和Bayesian取样。有关这些复杂设计概念和操作的阐述在其他书中可以找到（Frankel, 1983; Sudman, 1983）。

总 结

本章为初级研究者提供了一个有关取样理论操作步骤的基本概述。它从介绍取样对科学研究和增加研究发现可信度的重要性开始，再区分非概率取样和概率取样，并对主要的实验和调查设计作了回顾。现在我们要提醒研究者、特别是样本设计者有关实验伦理的问题。这里有一些敏感的问题，要努力保证特定总体中的所有成员都有机会在实验样本中出现，与研究者和政策制定者共享其观点。还有一点，应该努力保证被试是自愿参与实验或做出反应，研究没有强迫或欺骗成分。

推 荐 阅 读

以下的读物仅仅是为期望自学和进一步提高的人提供一些文献实例。其实，没有什么可以代替向别人提问、自己设计取样框、向同行、导师征询意见、不断向自己发问等这些实践活动。取样的数学基础可以参见Cochran (1977)。另外一些关于取样的基础知识，可见Kalton (1983)、Kish (1965) 和Sudman (1976)。一个由Sudman (1983) 撰写的有关应用取样的章节列出了许多有用书籍，这些书籍由有经验的研究机构编撰，包括芝加哥大学国家观念研究中心 (NORC)、美国人口普查局以及密歇根大学社会研究所 (ISR) 下属“调查研究中心” (SRC) 等。有一份不错的小册子列出了取样的定义 (调查取样公司, 1995)。也可以通过互联网搜索引擎进行在线搜索。可以联系调查取样公司: info@ssisamples.com。与Genesys公司联系, 可拨打电话: (215) 653-7100。

参 考 文 献

- Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). New York: John Wiley.
- Frankel, M. (1983). Sampling theory. In P. Rossi, J. Wright, & A. Anderson (Eds.), *Handbook of survey research* (pp. 21-67). New York: Academic Press.
- Goodwin, C. J. (1995). *Research in psychology: Methods and design*. New York: John Wiley.
- Groves, R. M., & Kahn, R. L. (1979). *Surveys by telephone: A national comparison with personal interviews*. New York: Academic Press.
- Kalton, G. (1983). *Introduction to survey sampling*. Beverly Hills, CA: Sage.
- Kish, L. (1965). *Survey sampling*. New York: John Wiley.
- Mosteller, F., Rourke, R. E. K., & Thomas, G. B., Jr. (1961). *Probability with statistical applications*. Reading, MA: Addison-Wesley.
- Murray, J. (1978). *Continuous national sample: Project reports*. Chicago: National Opinion Research Center.
- Piekarski, L. (1994). *Random digit telephone sampling methodology*. Fairfield, CT: SSI.
- Skinner, B. F. (1966). Operant behavior. In W. K. Honig (Ed.), *Operant behavior: Areas of research and application* (p. 21). New York: Appleton-Century-Crofts.
- Sudman, S. (1976). *Applied sampling*. New York: Academic Press.
- Sudman, S. (1983). Applied sampling. In P. Rossi, J. Wright, & A. Anderson (Eds.), *Handbook of survey research* (pp. 145-194). New York: Academic Press.
- Survey Sampling, Inc. (1995). *Glossary handbook: A survey researcher's handbook of industry terminology and definitions*. Fairfield, CT: Author.



第三部分

收集数据



第 9 章

申请用人类被试做研究

Lyle D.Schmidt

Naomi M.Meara

用人类作为被试进行研究前，必须先得到审查委员会（Institutional Review Board, IRB）的批准。为了保证研究计划的顺利进行或按时完成任务（如完成学位论文等），我们必须清楚如何做这件事。为了快速而有效地得到批准，你不单要了解所在大学采用的具体申请程序和形式，也应理解建立IRB的理由，即其道德评价和审查权力的基础。

审查委员会的发展

在美国，只要研究需要人类被试，任何研究机构——无论是受联邦政府部门（或代理机构）管理、支持还是调控——都必须有保护研究被试的程序。1991年修订的联邦法规第45条CFR部分第46项，明确了管理此类程序的联邦政策。每个机构都要根据联邦政策制订出自己的指南，并和科研风险保护办公室（OPRR）协商出一份保证书（OPRR是“健康和人类服务部”DHHS中的一个单位，其职责是保护人类被试）。如果你的协会已经和OPRR协商出一个保证书，那么你就应该在“人类被试审查委员会”、“人类被试保护制度审查委员会”或者一些使用这种术语的协会目录中找到一个办事处。这个办事处应该能够提供为你申请人类被试委员会的批准所必需的指南和资料。实际上，你的同事和导师可能不仅掌握以上这些信息，而且具有丰富的申请经验。

在绝大多数涉及人类被试保护的政策法规中，人们使用审查委员会（Institutional Review Board, IRB）这个名词。IRB是一个行政机构，目的在于保护被试的权利和健康。根据要审查的研究领域，IRB可以有一个或更多的部门，例如，社会和行为科学审查委员会或生物医学科学审查委员会。这些委员会审查你（研究者）的研究草案。这些草案描述研究项目或者个人调查，它们提供了详细的资料，以供委员会评估研究中人类被试的活动性质，确定研究者是否为保护被试的权利和健康设计了足够的措施。

为什么要建立IRB？尽管人们关注接受专业服务的人的权利已有一个很长的历史，例

如医生的希波克拉底誓言。但在美国,对人类被试进行保护的历史始于1947年。在纳粹集中营中,医生和科学家们对囚犯滥用生物医学试验,这些罪行在1945年纳粹德国战败后被公布于众。纽伦堡战争罪行审讯期间,纽伦堡法案诞生,并被用来作为判决此类罪行(OPRR, 1993, p.A6-9)的标准。这个法案在原则上阐述了对待参与研究的人类被试的道德要求,并成为以后法案的典范。最初的原则中“人类被试的自愿同意”是最核心的要求(OPRR, 1993, p. A6-1)。其他原则包括:良好的风险受益比,避免伤害和患病,保护被试免受伤害,保证研究人员是合格的科学家,保证被试有随时退出的自由。1964年制定的世界医学协会赫尔辛基宣言及其1975年、1983年和1989年的修订版(OPRR, 1993, p.A6-3)均对这些原则做了详细阐述。纽伦堡法案和赫尔辛基宣言都是关于医学研究的。

在美国,保护人类被试的政府法规首次出现于1966年,即“国家健康研究院”(NIH)设立的保护人类被试条例。1974年,《国家研究法案》的通过,创建了“生物医学和行为研究中保护人类被试国家委员会”。1978年,该委员会提交的最终报告,题为《贝尔蒙特报告:对研究中人类被试保护的道德原则和指南》(*The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*)。该报告名称来自于委员会在Smithsonian 协会的Belmont会议中心开会的地点。此报告成为OPRR 1993年出版的《保护人类研究被试》(*Protecting Human research Subjects*)的附录6。

Belmont报告中对研究和实践进行了区分,并确定了在人类被试研究中的三个基本道德原则。报告中将研究定义为“一个被设计用来检验假设、能够得出结论,并且因此发展或促进知识总结概括的活动”(国家委员会, 1979, p. 3)。而实践的定义是“仅仅被设计用来提高个体患者或顾客的健康、且有一个合理的成功期望值的干预活动”(国家委员会, 1979, p.3)。委员会论述的三个基本道德原则是尊重、仁慈和公正。尊重涉及两种信念:人应该被作为一个独立的个体来看待,并且随着人身自由的消减,人们应该受到保护。仁慈是指,通过尽可能大的益处和尽可能小的损害来保护被试。公正是关于公平地分配研究的负荷与利益。

作为对Belmont报告的回应,1981年,美国“健康和人类服务部”(DHHS)修正了其关于人类被试的规章制度。1981年,这些制度被编入联邦法规全书的第45章第46部分“保护人类被试”(45 CFR 46)。修订本于1983年和1991年生效,后者涉及保护人类被试的联邦政策,即通过联邦政府的16个办事处来支持、运行或规范人类被试研究。由这些代理机构执行这些规则,适用于“所有采用、支持或者在其他方面符合这些规则的、涉及人类被试的研究”(45 CFR 46, 1991, p. 4)。

那么, Belmont报告中所描述的三个基本道德原则在研究上的要求到底是什么呢?第一条是尊重人,这是知情同意的基础。在被试能力所及的范围内,必须允许他们决定什么可以和什么不可以发生在他们身上。这个过程包括知情、理解和自愿三个因素。知

情的具体要求包括澄清研究程序、目的、风险和可能的受益；告知治疗中其他可供选择的方法；提供被试提问的机会；保证被试在任何时候都可以退出研究。因为不可能保证让被试了解所有相关信息，所以知情的标准是要告知被试那些看起来重要、且足以影响被试最终决定是否参与研究的信息。这种信息包括他们要做什么，他们的数据如何被处理，而且，如果数据不是匿名的，保密的恰当方法是什么。理解是指信息需要以适当的方式传达，以便每个被试都明白他或她将要面对什么。理解力有限的人（如儿童），可能需要第三方代理。即便如此，也必须保证他们自己有权选择是否参与研究，除非该研究的治疗有成功的可能，且没有其他可行的办法。自愿则要求在没有强迫或不当影响的条件下参与研究——所谓不当影响是指，例如，给被试过多的或无理由的报酬。

第二个基本道德原则是仁慈，这是风险/受益比分析的基础。应该把被试受到伤害的可能性和严重程度降低到最小程度，并且权衡它们对他人健康及福利所引起的利弊。这个分析涉及对风险和受益的性质和范围的考虑以及对它们的综合评估。这就给研究者提供了一个评价研究设计的机会，并可以帮助人们决定是否愿意参与研究。在下列的考虑事项中有几个是必需的：

1. 从道义上来讲，决不允许残忍或无人道地对待人类被试。
2. 风险应该仅限于必需水平，并且在研究设计中，应该尽量寻找降低风险的备选方案。
3. 必须阐明选择易感人群的合理性。
4. 在获得同意的过程中风险和受益必须表达得非常清楚。

第三个基本道德原则是公正，它是公平选择被试的根据，即平等地分配研究负担和受益的根据。就个人而言，不能仅对偏爱的被试提供可能有益的研究，而对不偏爱的被试提供可能危险的研究。就社会而言，公正可能有一个先后次序，如成人先于儿童；可能的情况下，公正要求回避那些已经为他们的条件或环境所累、处于弱勢的群体，例如被收容的人。如果主要以方便为前提选择被试，不公正就可能发生，例如少数民族、重病患者或者学生人群，他们碰巧在研究实施的地点，或他们的自身条件使其更容易被控制或更容易同意参加研究。公正通常就是说，那些可能从研究中受益的群体也应该是那些接近参与研究的群体。然而，有些人会认为，对于能够使他人获益的研究，尤其是当不具有危险时，个体有义务为社会利益去参与研究。

最有名的实施社会和行为研究的早期法规是《人类被试研究中的道德原则》（*Ethical Principles in the Conduct of Research With Human Participants*），最早是由美国心理学会（APA）在1973年出版，并于1982年修订（APA, 1973, 1982）。Milgram的工作（1965, 1974）和其他一些研究中，对所用欺骗的恰当性、知情同意以及对参与者的有害影响都存在疑问（对此类研究的一个批评，参见Bok, 1989），在这些研究之后，对社会科学研究的专门性或大众性评论成为这些标准产生的源动力。这些批评促使APA针对

研究道德进行了一个全面的回顾。与制订最初的APA道德规范法规（Hobbs, 1948）所用的程序相类似，判定新标准也采用了经验的方法。APA的成员接受调查，并描述那些涉及道德问题的研究。一共收到了约5 000个实验道德问题的描述。把这些事例及其他来源的描述转发给APA成员，请他们审查，然后再进一步修改和散发。在这个过程中发展形成的原则成为后来出版和修订的基础。

然而，应当指出，首次关于APA道德规范法规的出版物《心理学家的道德标准》（*Ethical Standards of Psychologists*, APA, 1953），已经包含了研究的标准。

这些道德标准包括心理学研究者对充分地计划和实施研究、报告研究结果和处理与被试关系的责任。具体原则包括：通过免除被试人员经受不必要的情绪压力，以及去除任何可能发生的有害副作用等方式，来保护被试的健康；当存在发生严重副作用的危险时，要让被试充分知情，并且给他们拒绝参加研究的机会；只有当研究的问题有明确的要求，且遵守了上述原则的情况下，才可以对被试隐瞒信息或给他们提供错误的信息；没有明确的许可，不能暴露被试的身份；提供承诺给被试的义务，感谢被试在研究中的合作（Schmidt & Meara, 1984, p.66）。

后来的APA道德规范法规修订本保留了专门针对研究方面的章节，并且包含这些普遍性原则。目前应用的法规《心理学家的道德原则和行为法规》（*Ethical Standards of Psychologists and Code of Conduct*, APA, 1992），包含有关人类被试参与研究的14个具体标准。这些标准含有与1953年法规（APA, 1953）中许多相同的主题，例如，知情同意、欺骗和履行承诺。同样，这些章节与在Belmont报告中所提出的普遍性原则（尊重、仁慈和公正）是相关的，并与IRB对研究回顾中的关键问题（讨论如下）有关。除了这些标准，目前的APA法规（APA, 1992）中有几项关于研究者诚实的标准，这些话题包括剽窃、出版信用和审阅别人所发表作品的专家的责任感。同样值得注意的是，在1992年法规中有一个直接涉及审查核定的具体标准（Standard 6.09）：“在实施研究之前，心理学家必须得到主办机构或者组织的恰当批准，还需要提供有关他们研究申请的准确资料，并且必须按照批准的研究草案来实施研究（p. 1608）。”

IRB审查的关键点

IRB对你草案审查的关键点可能是什么？它们是依据上述的研究要求和基本的道德原则得出的，并且在IRB（OPRR, 1993）中有详细的讨论。

知情同意

第一个关键问题是知情同意。对于审查研究草案的人来说，知情、理解和自愿的原

则是很重要的因素。研究的描述是否足够详细，是否保证被试知道什么可能会在他们身上发生，以及描述的方式是否能让被试理解描述的含义？邀请被试参加研究的环境是否可以让他们自由地做出决定？

国家对应该提供给入选被试的信息在联邦法规全书（45 CFR 46）、IRB指南（OPRR, 1993）和你所在协会的IRB政策和程序资料中都有所报告。审查者会审阅你的草案，并依据你所在的研究机构和审查者对这些草案的诠释，决定你是否已经达到了这些要求。他们可能特别关注如下几方面：同意书的内容是否充分？是否采用了容易理解的语言和形式？你是否已经准备了合适的同意书？如果你请求放弃书面同意，那么你是否已经提供了足够的理由，并且这些理由与联邦政策以及你所在的IRB的政策相一致？如果你的研究涉及现有的数据记录，或者打算在被试不知道的情况下对他们进行录像或拍摄，那么你是否已经达到了在这些情形中获取同意的特殊标准？如果你打算欺骗被试或者只告诉发生在他们身上的一部分事情，那么你是否已经提供了必须这样做的所有理由，而这些理由是否符合IRB的要求？

风险/受益分析

第二个关键问题是风险/受益分析。如果一个调查研究给被试带来危险，那么就应该通过对被试或社会的预期获益来证明他们的参与是合理的。风险来自于参与研究而对身体、心理、社会或者经济上产生损害的可能性。最小风险是指在研究中受到伤害的可能性不超过在日常生活中或者常规情况下身体或心理检查或测试中受到伤害的可能性。

一般说来，IRB成员都很关注研究加于被试身上的任何风险，如果可以带来益处的话，也关注研究可能给个人或社会带来的益处。正如Beauchamp和Childress（1994）所指出的：“在向IRB递交一份涉及人类被试研究草案的申请时，要求调查者罗列对被试的风险以及对被试和社会可能的利益，然后解释为什么利益会高于风险（p. 291）。”IRB的成员常常对风险和利益之间的平衡做一个理性的评估，而不是单纯依靠数字或统计。一个风险/利益分析涉及仁慈、尊重和公正的原则。例如，如果一个研究看起来没有潜在利益，那么这个研究很可能不会获得批准。有时IRB会指出，一个设计方案没有达到科学标准。尽管这样的研究可能没有任何重大风险，但其结果的可解释性可能非常不明确以至于没有明显的得益。设计拙劣的研究将会浪费被试的时间和被试人群资源。因此，这对于参与者和那些需要被试参与能更多获益的研究的学者来说，是不公平的。如果同意书不能体现出对被试恰当的尊重（例如，包含误导和不能清楚地描述潜在的风险和利益），IRB很可能会要求你修改申请书，以达到这个重要标准。在准备草案时，要求你不仅要向IRB成员们陈述风险和利益，而且必需证明你为被试提供的知情同意和其他信息怎样能使被试知情，从而做出是否参与的决定。

隐私/保密

IRB审查的第三个关键问题是隐私/保密。隐私与一个人的控制程度、时间选择或与他人共享的物质、行为或智力方面的环境有关。保密是指，未经过本人同意，不得把与被试之前约定以外的内容告知别人。

对一个对调查者来说，常规的、非侵害性的研究程序，可能会要求被试透露一些非常隐私的信息。被试经常被要求回答那些从未透露给别人或只对一两个好朋友说过的问题，因此隐私和保密的问题对于合理的研究以及你的研究计划是否被批准是非常关键的。

研究者和被试有着确切的约定，即被试按照他们所说的去做，并按照双方事先同意的方式处理获得的信息。虽然一个被试可能已经同意对其信息的使用，但是他（她）有权在任何时候从数据中除去可识别的信息，记住这一点很重要。通常，最好在数据档案（电子拷贝或者硬拷贝）中不要含有任何种类的、能够表明被试身份的标记。如果身份证明是必要的，那么研究必须采取额外的防范措施，除了那些在最初协议中详细说明的人以外，任何人都得不到这些数据。当发生违背协议的情况时，必须告知被试。目前，美国心理学协会法规（APA，1992）标准6.16规定：“对个人的、可识别的研究数据的进一步使用或者共享以及没有预期到的未来使用的可能性，心理学者都必须告知研究的参与者（p.1609）。”

APA法规（APA，1992）提供了关于侵犯被试隐私的规范。标准6.13增加了对被试拍摄或录音的特别警告，并且标准6.17要求调查者把侵犯降低到最小程度。除了把侵犯最小化，另一个要遵守的有益的综合规则就是，潜在被试有权不受打扰，并且不会被再三地要求参与研究。有些人认为，不论研究计划或者潜在的利益有多重要，研究者都不应该无视单纯的拒绝而纠缠一个潜在的被试。如果一个人说“不”，那么研究者就没有理由对计划的优点和个人的义务发动推销游说。如果说个体有帮助社会的义务，那么也不应该由研究者来“迫使”个体施行此责任，或者迫使个体决定谁“应该”参与一个特定的计划。研究者应该采用的最好态度似乎就是，如果人们同意作为研究被试，那么这些人实际是在做超越其义务的事。他们的动机可以是帮助研究者，贡献更多的善行，或者为他们自己赢得一些不大的利益，例如在研究过程中获得额外荣誉。不管动机是什么，潜在被试有拒绝的自由；当他们同意参与时，也应该仅仅是为了科学目的而利用他们所讲述的事情。不要轻视在研究中所获得的信息，这些信息不能被用来贬低被试或违背他们对研究事业的信任。

参与者的选择

第四个关键问题就是参与者的选择，IRB要考虑研究对于参与者在负担和利益上的公

正和平衡。参与的担子会落到那些最可能从研究中受益的群体中吗？如果可能的话，能否避免对某个参与群体的反复要求？研究对某个特定群体的选择有依据吗？如果准备使用易受伤害的被试，那么该研究能否改为在其他较不易受伤害的人身上实施？人类研究是很花时间的，并且经常很难去组织。物理实验室和化学试验或多或少都有固定的实验对象可以自由支配。人类难以组织，比起非生命物质的研究，心理学在关于权利和健康方面有更多需要考虑的事情。研究者往往倾向于选择那些容易集获的人群：大学生、学龄孩子、囚犯和其他州或联邦政府收容的人。这些贫穷的、未完成教育的或者垂死的人（在医学研究的情况下），往往不公平地承受着心理研究的负担，却并不能从中得到与之相应的益处。我们也意识到那些非主流文化的人比其他人更多地承担参与者的责任。分清公正与个人认为的、“理所当然”的科学和日常假设，对科学是很重要的。对这些问题的尊重和敏感是伦理科学的重要内容。在使用这些容易得到的参与者的过程中，我们应当尽可能防止对公正原则的违背。正如我们在下面要讨论的：当研究中涉及奖励因素时，这个问题可能会变得更加复杂。奖励应该公平分配，但不能强迫，尤其是对那些比较易受伤害的人群。

对参与研究的奖励

第五个关键问题是提供给被试参与的奖励。为了确保被试参与研究的决定是自愿的，IRB将试着测定被试的决定是否在没有强迫的和不合理的影响下所诱发的？征募时是否有微妙的或者明显的威胁？是否允诺参与者极高的报偿以使他们难以拒绝？

在尊重和感谢被试付出的时间和努力（并因此提供给被试一些报酬），与增大诱因以致结果受到不适当影响之间，存在一个精细的界线。例如，学生能够用其他方式赢得相同的学分，并且这种方式几乎投入同样多的时间和能量，那么在一个心理学班级里提供额外的学分似乎是很恰当的。要求学生通过参与研究来获得他们的很大部分学分（例如10%~30%），这可能是强迫性的。在这种情况下会很难证明被试同意参与确实是自愿的；它可能更应该被看成一种对知情同意原则的违反。目前的美国心理学协会标准（APA，1992）提供了对这些问题的指南。例如，标准6.14b规定：“心理学者不能提供过多的或者不适当的金钱上的或者其他的诱因来获得被试，尤其是当这种诱因对被试产生某种强迫性的时候（p.1609）。”

当研究者与潜在被试有特殊关系时，研究者必须谨慎从事。例如，教师和咨询师比学生和来访者更有权力，如果要从班级中或者来访者名单中招募被试的话，他们就必须对这个事实很敏感。这种权力差异能使潜在被试认为，除了参与，他们没有真正的选择权。一个人越有权力，他就越要小心尊重被试拒绝的权力。在为了一些预期的研究利益而改变来访者、咨询师之间关系的性质，还是忠于合约确定的治疗关系之间，这两者应该保持一个平衡（Meara & Schmidt，1991）。美国心理学协会标准（APA，1992）对此

提供了一些帮助。标准1.18对这种交换提出了忠告和指导建议；标准6.14a对以提供心理咨询服务来作为被试参与研究的奖励的心理学家所应承担的责任给出了专门的要求。对于大部分研究生来说，交换可能并不是一个问题，因为他们通常没有提供额外学分或者延长心理服务的机会。然而，标准6.11c（APA，1992）很值得注意，由于研究生常常做助教工作，需要慎重考虑学生是如何理解老师的要求和建议的。标准如下：“当心理学家用诸如学生或下属这样的个体进行研究时，要特别注意保护潜在的被试不会受到由于拒绝或者从参与中退出而产生的不利影响（p. 1608）。”IRB成员需要确定潜在的被试在这方面得到保护。简言之，IRB成员将会详细了解研究方案，以便确信研究的程序和防范措施不会以或明或暗的方式剥削和强迫被试。

为审查委员会（IRB）的审查作准备

在准备IRB的申请时，应该做的第一件事就是提前计划。如果没有这样的计划，你常会发现，虽然有可用的被试，而且项目已经作好收集数据的准备，但是申请却还没有获得批准。IRB通常约一个月召开一次例会。你需要知道他们何时碰头以及需要在会议前多长时间拿到你的计划。知道这些事情对于规划你自己的工作是非常重要的。同样，如果你是一个学生，那么了解自己导师的政策也是非常有用的。大学的IRB要求教员应是学生研究计划中的一个参与者。因此，你得知道你的导师需要多长时间阅读你的计划，而且你要用多长时间去修改这个计划。

你得清楚自己单位的IRB办公室的位置，这样你才能拿到申请表格。在本章最后的附录9.1中复制了一些这种表格的例子。填完这些表格后，你要请你的资助者或导师来审查。要确保有时间去修改。如果你第一次填表，可以先看一下已被批准的类似的研究领域的协议，这会对你很有帮助。学习一个已经成功的人可以提醒你，别人是如何处理上述我们讨论的关键问题的，并且会给你一些关于用不同的方式在一个知情同意的文件中表达重要问题或描述研究工具的信息。通常你的指导老师、其他教员或者学生可能愿意共享已经被批准的协议实例。

在上交协议时，你可能会询问在你的方案被评审的时候，是否可以参加。如果不可以，那么可以在完成自己的协议前，但必须在熟悉了表格的要求后，参加一个IRB会议，这会很有帮助。

一个人在完成研究协议过程中的态度是很重要的。通常，学生，包括一些教员，都把IRB看成是一个敌手或者最多也就是一个浪费时间而又不得不克服的障碍。其实IRB成员的要求在很大程度上与你的期望是一致的：一项符合伦理的成功的研究。通过了解自己的学科和研究的问题、方法、设计，可以学习政府、你所在研究机构、APA认为是重要的伦理原则，你可以试着与IRB达成一致。IRB成员把他们许多处理建议和协商程序的

经验添加在表格中，有关建议的优点和潜在问题的观点可以在这个表格中得到交流。详细计划、关注细节、模仿他人的工作、观察IRB的审议过程、了解自己与IRB的目标的一致性，能够促进你的方案的批准。这些步骤可以使批准过程变得有效而有教育意义，甚至很有乐趣。

推 荐 阅 读

尽管最初的APA《人类被试研究中的道德原则》（*Ethical Principles in the Conduct of Research With Human Participants*，1982）和1981年APA道德规定的第九条原则是相同的，但是这本书目前仍然有一定价值。它详细探讨了有关进行人类研究的几个问题，例如在研究关系中去除不当的利用关系、获得知情同意的义务和确保参与者免受强制的压力。APA可能很快会发布一个修订本。尽管这个修订本所包含的原则不会改变，但这些原则可能与目前APA法规中的标准6.06到标准6.19有直接关系（APA，1992）。

在“欺骗性的社会科学研究”一章里，《谎言：公共和私人生活中的道德选择》（*Lying: Moral Choice in Public and Private Life*）一书的作者、哲学家Sissela Bok（1989），批判性地审查了在社会和行为科学研究中欺骗的使用。她尤其强调了伪装和掩饰谎话如何干扰潜在参与者的选择权、危害研究者和参与者、且减少公众对研究的信任。

在《心理学科学》（*Psychological Science*）杂志中专门有一部分“研究中的道德”（1994，Vol. 5），是由Robert Rosenthal（1994）的一篇论文和Pomerantz、Parkinson、Gurman和Mann4人的评论所构成的一个综合文章。Rosenthal认为，科学和道德是紧密相连的，高质量的科学研究比质量较低的研究在科学伦理上更站得住脚。Pomerantz总的来说是赞成Rosenthal的，但是他认为，Rosenthal提出的用成本—效用分析来决定研究的道德可能太苛刻了。这种分析可能会使那些低水平或犯了一个简单错误的研究者容易遭到道德上的苛责。Parkinson不同意Rosenthal对科学和道德关系的解释，他认为，“一个研究的科学性独立于它的道德性。”Gurman提出，现在缺少对那些根据最初的问卷回答或其他筛选程序而从研究中排除的被试，应该有一定的事后解释，让其知情。最后，Mann介绍了一个用大学研究生进行的研究，评估了研究生对知情同意的理解程度。结果表明，在阅读和填写那些表格之后，被试对（表格中所包含的）实验的关键方面，或当他们同意自己的一些权利并不十分了解。Mann认为，即使不能采用口头的同意程序，也应该用较短的、条文不太冗繁的知情同意表格。

参 考 文 献

- American Psychological Association. (1953). *Ethical standards of psychologists*. Washington, DC: Author.
- American Psychological Association. (1973). *Ethical principles in the conduct of research with human participants*. Washington, DC: Author.
- American Psychological Association. (1982). *Ethical principles in the conduct of research with human participants* (Rev. ed.). Washington, DC: Author.
- American Psychological Association. (1992). Ethical principles of psychologists and code of conduct. *American Psychologist*, 47, 1597-1611.
- Beauchamp, T. L., & Childress, J. F. (1994). *Principles of biomedical ethics* (3rd ed.). New York: Oxford University Press.
- Bok, S. (1989). *Lying: Moral choice in public and private life* (2nd. ed.). New York: Vintage.
- Code of Federal Regulations, Title 45 CFR Part 46. (1981, 1983, 1991). Washington, DC: Department of Health and Human Services, National Institutes of Health, Office for Protection from Research Risks.
- Gurman, E. B. (1994). Debriefing for all concerned: Ethical treatment of human subjects. *Psychological Science*, 5, 139.
- Hobbs, N. (1948). The development of a code of ethical standards for psychology. *American Psychologist*, 3, 80-84.
- Mann, T. (1994). Informed consent for psychological research: Do subjects comprehend consent forms and understand their legal rights? *Psychological Science*, 5, 140-143.
- Meara, N. M., & Schmidt, L. D. (1991). The ethics of researching counseling/psychotherapy processes. In C. E. Watkins & L. J. Schneider (Eds.), *Research in counseling* (pp. 237-259). Hillsdale, NJ: Lawrence Erlbaum.
- Milgram, S. (1965). Some conditions of obedience and disobedience to authority. *Human Relations*, 18, 57-75.
- Milgram, S. (1974). *Obedience to authority*. New York: Harper & Row.
- National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). *The Belmont Report: Ethical principles and guidelines for the protection of human subjects of research*. (Reprinted in Office for Protection from Research Risks, 1993, *Protecting human research subjects*, Appendix 6, pp. A6-7 to A6-14)
- Office for Protection From Research Risks, U.S. Department of Health and Human Services. (1993). *Protecting human research subjects: Institutional review board guidebook* (2nd ed.). Washington, DC: Author.
- Parkinson, S. (1994). Scientific or ethical quality. *Psychological Science*, 5, 137-138.
- Pomerantz, J. R. (1994). On criteria for ethics in science. *Psychological Science*, 5, 135-136.
- Rosenthal, R. (1994). Science and ethics in conducting, analyzing, and reporting psychological research. *Psychological Science*, 5, 127-134.
- Schmidt, L. D., & Meara, N. M. (1984). Ethical, legal and professional issues in counseling psychology. In S. D. Brown & R. W. Lent (Eds.), *Handbook of counseling psychology* (pp. 56-96). New York: John Wiley.



附录9.1 俄亥俄州州立大学人类被试审查委员会行为和社会科学研究概览表及调查者核对单

“研究风险办公室”收到的超过最后期限（会前周五的中午）的方案，将被安排在下次会议上讨论。如果所有时间都安排满了，并且是在最后期限前或当时收到的方案，将被安排在下次会议上讨论。会议只评审已经完成的方案。填写不完全的方案将被退回。

主要调查者：

（如果是学生，列出
导师的名字）

打印名

签名

打印名

签名

打印名

签名

学术头衔：_____ 电话：_____

大学：_____ 系：_____

学校地址：_____

（教员的学校地址）

方案名称（如果有和方案不同的提议名，请用括号注明）：

计划研究的资金来源（选择A或B）

A. OSURF： 赞助者_____ RF提案/项目号_____

B. 其他 _____

有关人类被试研究活动的资金/赞助的信息仅为管理需要，在审查过程中，这种信息通常不作为考虑条件。

学术资源 PDG

观察记录表

以下的概览需同你的提案一同上交。要非常明确地描述被试在参与你的研究时会遇到的事情以及保护他们的措施。仔细对待以下栏目，将会有助于审查过程。

1. 用一两句话描述研究的背景和目的。
2. 简单描述在研究中所包含的每种条件或操作。
3. 在研究里会采取哪些操作或观察？如果要使用任何问卷、测试或其他的方式，请提供一个简单的描述，并且要包括一个副本，或者要说明这个副本何时能送来评审。
4. 被试会遇到心理、社会、身体或法律上的风险吗？会或不会，请描述。
5. 是否考虑到被试可能承受的压力？是或不是，请描述。
6. 会以任何方式欺骗或误导被试吗？是或不是，请描述并包含报告的要点或原本。
7. 要求被试提供那些他们认为是私人的或敏感的信息吗？是或不是，请描述。
8. 会给被试呈现一些他们可能认为具有攻击性、威胁性或侮辱性的材料吗？会或不会，请描述。
9. 每个被试大约会需要多长时间？
10. 谁将是研究的被试？如何征募或联系他们？必须告知被试作为参与者的性质，尤其要描述他们可能认为是不愉快或冒险的事情，请在被试自愿参加之前提供给予他们信息的要点或原本，包括征求被试所用的书面材料的副本和口头请求的要点。
11. 将采取哪些措施确保每个被试的参加都是自愿的？如果有，将会提供什么样收益的来让他们参加？
12. 你如何明确让被试在参加之前表示赞成，会用到一个书面的同意表格吗？是或不是，如果需要，请包含表格；如果不要，请说明原因。
13. 所有数据会成为确定被试身份的永久记录吗？有或没有？
14. 对于被试是否参与某个实验或研究的事实，会成为永久记录的一部分提供给导师、老师或雇主吗？是或不是。
15. 会采取什么措施来确保数据的机密性？
16. 如果研究中有任何危险，那么会不会产生对被试或社会有益的东西？
17. 是否会用到文件或档案资料中的任何数据？是或不是。



俄亥俄州州立大学行为和社会科学的人类被试审查委员会

检查清单*

供主要研究者使用——无需上交

- ____ 1. 是否包含研究计划的副本？它写得是否足够详细，以使委员会能够做出明智的决定？
- ____ 2. 是否包括详细目录/问卷的副本？
- ____ 3. 如果从其他外部机构（如学校、刑事机构等）征募被试，是否具有与这些机构达成的书面协议？
- ____ 4. 当征募被试时，主要研究者是否尽一切可能确保这些被试不是因为某些同辈间的压力而被利诱参加研究？
- ____ 5. 是否具有在获得被试同意前所应该告诉他们的内容的原稿？这些原稿是否写得与被试的理解能力相符？
- ____ 6. 主要研究者是否认为心理学的100 SING-UP程序可以满足书面同意的需要？
- ____ 7. 如果信件要寄给准备参与研究的孩子的父母，那么，这些信件是否包括以下内容：
- ____ 是否清楚地说明了研究是在俄亥俄州立大学进行的？
- ____ 是否告知了主要研究者的姓名？
- ____ 信中使用的语言能否被那些家长正确地理解？
- ____ 8. 主要研究人可否在需要的时候为那些书面同意的弃权者提供解释？
- ____ 9. 主要研究人是否采用了恰当的同意书（符合HS-027 or HS-028A标准），或者为了符合被试的理解水平而做了修改？需满足以下条件：
- ____ 同意书是否包括主要研究人的姓名、研究的题目、主要研究人的签名。
- ____ 当被试中有孩子时，是否征询家长的同意？
- ____ 10. 主要研究人是否区分了被试的匿名性和保密性？
- ____ 11. 如果在研究过程中进行了录音和录像，是否告知了被试或者其家长？
- ____ 12. 是否有被试情况的报告综述？这对包含欺骗的研究常常是必需的。
- ____ 13. 研究中是否使用了任何永久的记录？
- ____ 主要研究人是否可以提供足够的证据来说明信息获得的正确途径？
- ____ 是否告知被试可以查阅这些记录？
- ____ 14. 是否所有数据都将成为被试的永久记录？
- ____ 15. 是否告知了被试完成数据收集的精确时间？
- ____ 16. 如果在数据收集过程中发现被试需要立即进行心理上或生理上的帮助，对此如何操作是否也包括在内？
- ____ 17. 如果存在任何可能的医学风险，主要研究人是否已和具备资格的医生一起讨论过研究过程？
- ____ 18. 在适当的情况下，是否告知被试的私人医生，被试将参与研究？
- ____ 19. 当被试需要交通运输工具时，主要研究人是否做好了准备，承担由于意外造成的责任？

* 欲知在研究中使用儿童或因犯的指导原则，请致电研究风险办公室。

第 10 章

邮 件 调 查

Alan Vaux

在社会科学领域，邮件调查是一种得到广泛使用、适用性极强的方法。大部分研究型心理学家在职业生涯中都会搞一些调查，而且调查也是他们收集数据的主要方法之一。随着个人电脑、高品质打印机和高级软件的出现，进行高质量的调查已比以往容易得多。但是关于调查方法的正规培训内容，在心理学研究生的课程中仍相当少见。

本章旨在提供进行邮件调查的入门知识。我将针对正在采用（或者正在考虑采用）邮件调查法的研究者，讨论一些他们需要注意的关键问题，并提出有效进行邮件调查的建议。一位助研、一位初涉调查法或是一位资深研究者将从这里获得足够信息，使调查避免严重错误及满足一般科学标准。实际上，同其他所有研究方法一样，做好这种研究需要经验、创造力、智慧和不断的研习。有参考价值的文献见Dillman（1978）、Sudman and Bradburn（1982）、Fowler（1993）、Fink and Kosecoff（1985）以及Goddard and Villanova（本书第7章）。

调查法的优缺点

与所有其他方法一样，一个技术上无懈可击的调查未必能给知识带来重大贡献，但一项杰出的研究却可能因为采用了糟糕的调查而被削弱价值。调查方法可以通过内部效度、结构效度、统计结论效度和外部效度（Cook & Campbell, 1979）进行衡量。

内部效度

在内部效度方面，邮件调查可以与真正的实验一样有价值，在这些实验中，被试被随机地分配到不同的处理中。这种方法在大规模的现场试验研究中更有用处——那些现场研究以大样本去接受复杂的、不断进行的干预为特征。真正的试验研究当然也可完全通过调查进行，但这样的话，操纵变量、随机指派被试的余地就很有限了，作出因果关

系判断的能力也就受到了限制。无论是真正的试验研究还是现场研究，不同的应答率可能导致对选择性的威胁（例如，不同组之间不对等或无法保持对等），以至降低了内部效度，将研究削弱为准试验研究。总的来说，调查大多用在描述性研究、被动观察研究或者其他一些内部效度无关紧要、次要或没有其他更好方法的场合。真正的实验研究通常要求采用比邮件调查更为严格的数据收集方法。

统计结论效度

用统计方法发现经验关系，其效力有赖于影响统计能力的多种因素（样本大小、检验类型等），但是此处最相关的是测量的信度。邮件调查本身在信度方面并不低于其他方法。

要注意工具的选择以及对其信度的评价。为限制调查的长度而采用较短的量表会使信度受到威胁。更为常见的是：初涉调查的研究者往往忘记了信度的基本原则，而依赖单一项目去测量一个变量，尽管多项目可行且更好（记住木匠的格言：多用墨斗，省却锯头）。简短地说，标准的程序（仔细和清晰地用词）、多个准重复项目以及其他提高测量信度的方法，在邮件式研究中与在其他研究方法中一样重要和可行。

结构效度

关于结构效度，调查与其他种类的自我报告和他人报告的方法具有同样的优势和劣势。对许多心理学家关心的结构来说，比如信仰、人格、态度、价值、兴趣、行动，自我评估是重要的测量手段。他人，比如教师对学生的评估、主管对员工的评估、临床医生对病人的评估等，也是有价值的。其他一些结构，如血压或诱发电位，是无法通过自我评估或他人评估的方式进行测量的。

调查测量的信度与效度和其他测量方法一样可以检验。一些研究结果表明，调查甚至对诸如临床问题之类的“敏感”结构也能测量（Fournier & Kovess, 1993）。研究者要问的关键问题是，是否可以期望某个人提供自己感兴趣的信息？针对某个变量的有关的、可信的、有效的调查量表是否已经存在？一个现存的自我评估式量表是否可以转用于邮件调查之中？

外部效度

外部效度的一个主要方面与被试有关：我们可以将结论推广到哪些人身上？与邮件调查法密切相关的是涉及一般性和代表性等问题的抽样程序。这些问题经常为心理学研究所忽视。对期刊论文的检阅可以发现，虽然被试样本的大小和人口统计构成通常会被报告出来，但他们往往属于自愿者样本。一般情况下，文章很少报告征募被试的程序。

邮件与其他调查方法的对比

另一个问题是如何在邮件调查、电话调查、个人访谈以及电子邮件调查等方法之间进行选择。这些方法各有其优缺点，适用于不同的、特定的情况。有几种情况不宜采用邮件调查，这些情况包括数据收集比较紧急、问卷中开放性的问题较多、问卷采用复杂的依存设计（就是说，下一个问题是根据被试对上一个问题的回答而决定的）。在这些情形下，一方面建议采用电话调查、计算机辅助调查或个人访谈。另一方面，如果时间允许研究者仔细地寻找或者设计一套封闭式问题，并且数据能够在短时间内从一个相当大的样本中收集到，与电话调查相比，邮件调查可以节省开支获得质量较好的数据。

抽样和返回率

抽样

糟糕的抽样会破坏调查研究的外部效度，但与之有关的问题却常被忽视或误解。同其他方法一样，邮件调查可以针对方便的样本进行。然而在研究中，样本的代表性往往是很重要的。其实，正式抽样也可以较为方便地使用在邮件调查上。作为本章简短论述的补充，可参考其他作者如McCready（本书第8章），Cook and Campbell（1979）或者Henry（1990）等人的文献。对于许多研究目的，能代表研究者感兴趣的总体取样框是可以得到的。这些框架可以是组织良好的通讯录、专业名录、社区电话号码簿，或者从一些不同的数据库中综合产生的名单。正式取样可以是纯随机的、系统的，或者是更为复杂的——例如分层式随机抽样或聚类抽样。随机抽样意味着给个体或者抽样单位编号（也就是说，每个个体或者单元在抽样框中被选择的可能性均等）。系统抽样是在每第 n 个个体或单位中抽取一个（ n 是抽样框与期望抽取的样本大小之间的比例）。周期性随机选取开始点有助于确保每个第 n 个个体不是与某种特殊类型巧合，如组织者或者住在公寓拐角的人。分层随机抽样是在一定的层面进行抽样。例如，随机地从定义为性别、心理协会分会以及国家不同地区的不同组别里进行抽样。聚类抽样是在不同的层面进行，例如，随机地选取一些州，然后大学，然后系，然后教授。在不同的研究情况下，每一种方法自有其长处和短处。总的来说，正式抽样的优势依赖于调查的返回率。

返回率

可以采用许多办法提高返回率。这里需要指出的是，在计划邮件调查时（如申请课题或博士论文课题），重点应放在方法上，而不是结果上。也就是说，要特别注意确保能够获得足够大小的样本。最为重要的，关于返回率的研究文献越来越多，应该加以借鉴，以便斟酌在一定的研究条件下所需费用及收益。最起码，一份经过仔细设计的问卷

加上一封不错的附信通过批量邮件发出去，回执信封上应贴足返程邮资。两个星期后向每个被试发一张明信片，对填写了问卷的人表示感谢，对尚未返回问卷的人表示催请。除这些最基本的做法之外，Dillman（1978）和其他研究者建议，采用优先投递邮件而不是批量邮件以及许多其他特殊的方式，例如在每张附信上签名。其他一些可供考虑的方法包括：采用挂号信件（Gitelson & Drogin, 1992），进行多次的后继提醒（通过明信片或电话；Nederhof, 1988），向未曾返回问卷的被试重复投寄问卷，作预先通知（通过信件或电话；Schlegelmilch & Diamantopoulos, 1991）以及采用报酬激励（Brennan, 1992）。这些方法的可行性和开支可能会随着样本大小而变，但是研究结果表明，这些方法可以明显地提高返回率，至少在某些情况下是如此。Altschuld 和他的同事们曾经获得过高达96%的返回率（Altschuld et al., 1992）。

附 信

新手常犯的一个错误是对附信不在意，往往到事后才意识到这一点。Dillman（1978）提醒研究者们，这种沟通方式——包括得体的暗示——的重要性在于，他们是在请求一个陌生人为他们付出时间和努力去完成一项工作，它可能涉及这个人的感受、思想或者信仰。因此，除了保持礼貌和专业化，这封信必须起到如下的几个作用。

研究的重要性

首先，读者可能会问自己：“这是什么？我为什么应该予以理会？”因此附信的第一段就必须向被调查者介绍你的研究内容，并且指出它的重要性、可能的益处以及与被调查者的密切关系。根据你的研究课题，这可以写得比较概括，例如，“在美国许多社区中，暴力是一个严重的社会问题。这项研究的目的是帮助我们了解它，并为学校和公安部门提供如何在您所生活的城市控制暴力的方法。”或者能够引起更为直接的兴趣，例如，“行政介入正在改变着心理治疗业的营业方式和付费方式。作为一个独立开业的临床心理学家，你可能很关心这些变化。这项研究将有助于贵州的专业协会代表你们的利益去影响正在进行的立法。”诸如此类的例子皆以简短为要。参见Dillman（1978）和其他参考文献。

应答者的重要性

其次，收到调查附信的人可能会问：“为什么要我来参加？”附信要从两个方面回答这个问题：（1）解释为什么潜在的应答者会被选中；（2）他/她作为应答者的重要性。在参与者只是偶然被选中的情况下，这可能需要一些微妙的表达手法。例如：

您代表贵城一个相对人数较少的居民类型，被随机抽中来参加本研究。为了能够得

到所涉及问题的全貌，您在这个问题上的观点对我们来说至关重要。因此，如果您能抽出15分钟时间完成所附调查，我们将不胜感激。

隐私

潜在的参与者可能会问：“谁将会知道我在调查中的回答？”在涉及敏感信息的调查中，这个问题尤为突出。附信必须告诉潜在的参与者，他们的隐私权能在多大程度内得到保护以及你如何做到这一点。匿名是最容易传达的：“这是一项匿名调查。将完成的问卷装入所附的回执信封寄回，不要在调查或者信封上写上你的名字。我们不能由答卷辨认出任何人，甚至不知道谁参与或未参与这项调查。保密性是个比较难以传达的信息，因为在这方面是研究者在寻求参与者的信任：

“您的调查数据将被完全保密。您的名字在任何时候都不会出现在问卷上或与问卷一起保存。在您的问卷上有一个识别号码，但是它仅用于将这份调查与明年将要发给你们的另外4份问卷联系起来。名字与识别号码的清单将被封存，并在本课题完成之后予以销毁。”

研究者必须意识到有关职业道德方面的争论，并且给予充分注意。暗地里根据问卷识别问卷的参与者或有违调查保密性的承诺，这些对信任的背离都是对职业道德的严重触犯。为了提高返回率而发出具有欺骗性的保密承诺在职业道德方面也不能自圆其说，并可能招致适得其反的结果（参见APA，1992）。

总而言之，附信相当重要。它可能是研究者与参与者的第一次沟通，产生关键性的第一印象。在不同的段落里，该信应该交代研究的重要性，指出收信者参与的重要性，解释保护个人隐私的措施。最后一段可以重申这些情况，并鼓励人们就所关心的问题来电垂询。研究者会发现，科研职业道德管理委员会要求研究者在附信中插入一些他们规定的标准段落，或者采用与本文建议相悖的标准格式，这包括关于风险的标准声明、关于自愿参加的提醒及委员会的授权、说明研究者学生身份的内容等。有时这类标准声明损害了附信良好的设计，甚至严重地削弱了一个调查良好的设计，同时又未提高对参与者的道德保护。研究者必须权衡利弊，与职业道德委员会多加协调和沟通。

总体格式

如Dillman（1978）所指出，调查设计应该反映一个观念，即研究者是在请求一个陌生人的帮助。它应当有助于而不是有碍于人们的参与。这是个明显的道理，然而新手们经常设计一些颇不简练、杂乱无章的调查。它们以模糊不清和不甚礼貌的方式收集信息，使应答者很难回答问题，感到爱莫能助。概括地说，问卷应该看起来专业、有趣、易于完成。功能很强的个人电脑、高级软件、高品质打印机已经极大地减少了制作专业

化问卷的难度，尽管这同时也提高了衡量可接受应答的标准。我在此处的一些建议来自 Dillman (1978) 的“整体设计法” (Total Design Method)。虽然这些建议显得卓有成效，它们仍应被看作是样例说明性的。重要的一点是，调查应遵循前后一致的设计风格，以助于应答者的参与。

版面设计

总体上，调查、版面、字体等应该看起来专业、清晰、易于完成。如果潜在的应答者觉得研究者未曾用心设计或完成问卷会劳心费神，他们就不太可能参加。欲使调查出色并易于完成，可以采用几个措施：把调查做成小册子，可以使其显得更为专业也更为短小，同时也减少了纸张和邮递费用。无需采用缩印法，使用能灵活控制字体的文字处理器，就能很容易地制作出小册子。对折的纸正反皆印，能产生4页问卷；两折的纸能产生8页问卷。只要排版正确，任何一个复印商店皆可进行复印、打印和折叠。有经验的研究者建议，封面只印研究标题或者徽标，封底保持空白或者印上最后一个开放式问题以及感谢声明。

应该尽力为调查建立清晰、一贯的风格，使其具有简洁的外观和充足的空白；将问题与回答清楚地区分开来，建立一套答题模式；每一部分由一个转折段落开始，把读者引导至主题（例如，下面一系列说法是关于……），进行提问（在何种程度上您同意或不同意这些说法），并且对怎样作答做出指导（例如，“根据下面方框中的数值，对于每一个说法，圈下一个与您的回答相同的数字”）。项目与应答选择应该清楚地区分开来，这可以通过在空间上采用不同字体、阴影、外框或其他方式达到。应该使用前后一致的回答问题的方式，可以是勾方框、圈数字或其他形式。圈数字的方式有利于数据的输入和编码。当许多问题有一套相同的回答方式的时候（例如，一个四点刻度），应该先提供一个参考数表，然后按照项目在左、数表在右的方式排列。当不同的问题有不同的应答方式时，Dillman (1978) 建议，将每个问题的选项垂直排列。这似乎浪费了空间，但“节省空间”的格式往往让人感到混乱和困惑。尽管采用不同字体的方式有助于将问题嵌入卷页，但应注意避免因采用不同字体和字形而导致的混乱。

问题的组织

问卷应该按照不同的主题分成不同的部分，不同部分之间有简短的转折说明，以便从一个部分向另一个部分转换。在附信之后，问卷应以一个比较吸引人的问题开始。对于一个特定的主题，格式相似或应答格式相似的问题应该放在一起。这会使参与者易于完成，或者让他们感到比较易于完成。针对人口统计信息的问题应该放在调查的最后，而不是开始。

上下文

在组织调查的各部分时，研究者应该意识到，上下文，尤其是先前的问题，可能影响参加者的回答。Tourangeau和Rasinski（1988）提出了回答调查问题的认知模式。这个模式包括4个部分：确定问题的意旨，回忆相关的信仰和感受，使用这些进行判断，然后根据判断来选择应答。他们认为，上下文的材料，例如，上一个问题，可能会激活一套信仰，为后来的判断设定一套标准，或产生一种保持一致或折中的内在压力。Tourangeau和其他研究者证明了与一个特定信仰有关的上下项目是如何增加应答者提取有关信仰（Tourangeau, Rasinski, & D'Andrade, 1991）和改变调查应答的（Ransinski & Tourangeau, 1991），这种影响尤其发生在当参与者对一个主题持有成熟但又自相矛盾的观点时（Tourangeau, Rasinski, Bradburn, & D'Andrade, 1989）。简单地说，上下文效应可能会有助于对问题的回答（通过唤起有关信仰），但也可能改变应答。在个人访谈或电话调查中，通过采用多种调查格式，或通过计算机随机显示问题，就可以比较容易抑制这种效应。在邮件调查中抑制上下文效应则相对困难一些。

问 题 设 计

心理学家们经常在邮件调查中使用或改编使用一些成熟的量表，但由于种种原因，调查往往涉及一些新的问题。这些问题必须经过认真设计。此处的介绍只能是简略的，建议读者进一步阅读其他文献（Dillman, 1978; Fowler, 1993）。需要指出的是，这里提出的建议也可以用在邮件调查之外的其他调查方法的问题设计中。类似的是，从其他方法获得的关于问题设计的知识也会有助于邮件调查。如今正在进行的大规模调查，例如“综合社会调查”（Davis & Smith, 1992）经常包含对措辞效应的研究，且已获得一些有用的信息。调查设计也越来越重视考虑参与者的认知过程：对问题的理解、对信息的回忆、判断的进行、对应答的组织等等（Fienberg & Tanur, 1989）。到目前为止，由这些途径得到的知识尚未给问题的设计归纳出一套简明的指导性原则。大体上说，所设计的问题应该礼貌、清晰、准确、不偏不倚且忠于特定的研究目的。

礼貌

再没有什么比使用苛求的、粗鲁的、不敬的、高高在上的提问方式更能迅速地毁掉一个调查了。常见的错误是采用提问方式的问题（例如，年龄____），不必要的介入性问题（例如，您年收入的确切数字是多少？\$_____），明显苛求的问题（例如，根据重要性的次序将下面25个主题进行排列），不可能精确回答的问题（例如，在过去的年月里你总共阅读过多少本杂志），一相情愿的问题（例如，当你在海外旅行的时

候……)，高高在上的问题（例如，向有经验的科学家问：“发表是一个复杂的过程。手稿应该送交审稿，且经常在发表前进行修改。如何才能……”）。总之，询问应该始终采用提问的形式，回答起来比较容易，无需先决经验或知识，以得体的方式获取必要的个人信息。

清晰和准确

问题应该简短、直接但不模糊。所有应答者对同一个问题的理解应该没有差异。复杂的问题可以拆开。常见的错误包括：不必要的润饰（例如，有些人认为……另一些人认为……你怎么想），问题冗赘（例如，您认为美国政府赤字过于巨大并且支出应该紧缩吗），双重否定（例如，当今的研究生训练未能有效地教会学生如何不犯基本的研究错误）。

使用合适的语言——简单、简短的而不是冗长、复杂的词汇。一方面，应注意避免使用参与者可能不熟识的缩语和略语、技术词汇以及生僻的表达方式；另一方面，一位在某个领域有专长的人可能将调查者对技术性词汇和常用缩略语的误用，看作是缺乏能力的表现或居高临下的态度。总之，应该在导言中解释清楚应答者可能不熟悉的术语和习语；每个缩写第一次出现时应该放在该用语全称之后的括号里。

为了测量精确，手头的研究问题应当清晰。过于模糊的问题是没有用处的（例如，心理治疗的质量高否）。精确意味着在研究者的要求与参与者的表达之间做出妥协，否则得到的将是伪精确（例如，确切地说，您去年看过几场电影）。也应注意避免提出一些需要应答者具备先决知识的问题，因为他们可能并不具备那些知识。

避免偏差

避免以诱导的方式提问（例如，你是否同意APA大量的介入性要求给研究生教育带来了不必要的困难）。应答的选项也应该保持平衡，无明显偏向（例如，为了加强心理学研究生培训，APA的要求是否应该略有增加、略有减少、大量减少或者消除）。

收集敏感信息

在心理学研究中，经常需要收集敏感信息以及在他人看来较为个人化和隐秘的资料。可以通过设计较为宽泛的答题选项而使问卷中的问题变得含蓄（例如，设置较为宽泛的个人收入范围），或者通过一系列问题逐步接近主题（例如，先询问应答者对社会上毒品问题的一般观点，其他人使用毒品的情况，然后问到应答者使用毒品替代物的情况）。要在想对敏感信息刨根问底的冲动与可能导致的参与者流失之间保持平衡。

总 结

总的来说,邮件调查是一种灵活的、有用的和被广泛采纳的方法。若使用得当,它可以提供一个相对便宜的途径,收集在地理位置和社会组织上遥远的分散的大样本数据。邮件调查可以用来收集敏感信息,在某些情况下可能收集到比采用电话调查和个人访谈更为出色的数据,且所需费用更低。正式的抽样——尽管心理学研究中较为罕见——可以方便地结合进邮件调查。不断增加的研究文献表明,可以通过一些方法提高返回率。较高的返回率和高质量的数据依赖于深思熟虑的附信、设计良好的版面和仔细斟酌的调查问题。认真的研究者可以很容易地掌握这种调查法的原理。本章末尾的附录10.1展示了进行这种调查所要考虑的问题和需要完成的工作。

推 荐 阅 读

我们建议读者参阅如下文献,以获取本章所覆盖的主题的进一步信息。包括:Dillman (1978)、Fink和Kosecoff (1985)、Fowler (1993)、Henry (1990)及Sudman和Bradburn (1982)。其他有关的信息可参考Fienberg和Tanur (1989)。

参 考 文 献

- Altschuld, J. W., Thomas, P. M., McCloskey, W. H., Smith, D. W., Wiesmann, W. W., & Lower, M. A. (1992). Mailed evaluation questionnaires: Replication of a 96% return rate procedure. *Evaluation and Program Planning*, 15, 239-246.
- American Psychological Association. (1992). Ethical principles of psychologists and code of conduct. *American Psychologist*, 47, 1597-1611.
- Brennan, M. (1992). The effect of monetary incentive on mail survey response rates: New data. *Journal of the Market Research Society*, 34, 173-177.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin.
- Davis, J. A., & Smith, T. W. (1992). *The NORC General Social Survey: A user's guide*. Newbury Park, CA: Sage.
- Dillman, D. A. (1978). *Mail and telephone surveys*. New York: John Wiley.
- Fink, A., & Kosecoff, J. (1985). *How to conduct surveys*. Newbury Park, CA: Sage.
- Fournier, L., & Kovess, V. (1993). A comparison of mail and telephone interview strategies for mental health surveys. *Canadian Journal of Psychiatry*, 38, 525-533.
- Fowler, F. J., Jr. (1993). *Survey research methods*. Beverly Hills, CA: Sage.
- Gitelson, R., & Drogin, E. B. (1992). An experiment on the efficacy of certified final mailing. *Journal of Leisure Research*, 24, 72-78.
- Henry, G. T. (1990). *Practical sampling*. Newbury Park, CA: Sage.
- McHorney, C., Kosinski, M., & Ware, J. E. (1994). Comparison of the costs and quality of norms for the SF-36 Health Survey collected by mail versus telephone interview: Results from a national survey. *Medical Care*, 32, 551-567.
- Nederhof, A. J. (1988). Effects of a final telephone reminder and questionnaire cover design in mail surveys. *Social Science Research*, 17, 353-361.
- Rasinski, K. A., & Tourangeau, R. (1991). Psychological aspects of judgments about the economy. *Political Psychology*, 12, 27-40.

- Schlegelmilch, B. D., & Diamantopoulos, A. (1991). Prenotification and mail survey response rates: A quantitative integration of the literature. *Journal of the Market Research Society*, 33, 243-255.
- Sudman, S., & Bradburn, N. M. (1982). *Asking questions: A practical guide to questionnaire design*. San Francisco: Jossey-Bass.
- Tourangeau, R., & Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin*, 103, 299-314.
- Tourangeau, R., Rasinski, K. A., Bradburn, N., & D'Andrade, R. (1989). Carryover effects in attitude surveys. *Public Opinion Quarterly*, 53, 495-524.
- Tourangeau, R., Rasinski, K. A., & D'Andrade, R. (1991). Attitude structure and belief accessibility. *Journal of Experimental Social Psychology*, 27, 46-75.



附录10.1 一份邮件调查清单

- _____ 本研究要探讨的主要问题是什么?
- _____ 邮件调查是否适于本研究设计(如操纵、时间)?
- _____ 通过自我评估或他人评估的方式能否测量主要的变量?

邮件调查或其他方法

- _____ 本研究是否涉及一个较大的样本?(样本越大,邮件调查就越比电话或深度访谈具有优越)
- _____ 是否有设计问卷的足够时间?(邮件调查需要仔细准备,然后复制起来较为容易。)
- _____ 问题是否以封闭式为主?
- _____ 调查是否包含一些前后依存的问题? 是否有较易理解的跳跃?

抽样

- _____ 考虑到问题、设计和统计能力,你需要多大的样本?
- _____ 能否获得合适的样本框?
- _____ 采用何种抽样方法(如随机抽样、系统抽样)?

返回率

- _____ 你准备采用何种方法来提高返回率?(如提前通知、奖励、跟踪提醒、重复调查)
- _____ 30%的返回率是否可以接受?

附信

- _____ 你是否已对附信仔细斟酌,提供了关于研究的重要性、参与者如何被选择以及如何保护隐私的信息?

组织

- _____ 你是否已经选择了可靠的、有效的方式对主要变量进行测量? 这些方式是否曾在以往的邮件式研究中使用过或较为方便地改编过来的?
- _____ 你是否将研究中最感兴趣的问题放在了调查的开始,而将有关人口统计信息的问题放在了最后?
- _____ 你是否考虑过在调查中采用其他组织材料的方式?
- _____ 是否存在问题秩序效应或者上下文效应?
- _____ 如果调查看起来太长,你能否采用较短但又可接受的量表?

版面

- ____ 是否已经考虑过选择何种形式（小册子等）？
- ____ 是否采用了一套连贯的设计风格（页边距、字体、字号等）？
- ____ 问题和答题区域是否在空间上和形式上有效地被区分开来？
- ____ 是否建立了一套简单一致的答题模式？

问题设计

- ____ 你是否已经对那些通常以其他格式存在的量表进行了合适的修改，以适应你的问卷？
- ____ 你是否对你自己新增的问题进行了仔细的斟酌？
- ____ 你是否已对问题的礼貌程度进行了复查？
- ____ 问题是否清楚且足够准确？
- ____ 有些问题或答题选项是否有偏见？
- ____ 你是否已对敏感问题进行复查并且考虑过替代方式？
- ____ 你是否已经请有经验的和无经验的应答者完成问卷且提供反馈意见？

在复印之前

我几乎从来没有看到一份完全正确无误的问卷！从小的打印错误到灾难性的错误（例如，应答选择表的选项印成从“完全同意”到“完全不同意”）皆可发生。

- ____ 你、你的同事和一两个校对者是否仔细核对了调查清单？



第 11 章

电 话 调 查

Peter Y. Chen

电话调查已经成为在态度调查和消费行为调查领域广为使用的方法。在1970年以前，这种方法因其方法论上的薄弱而受到批评，尤其是当时美国的电话覆盖率低（1963年大约为20%；Thornberry & Massey, 1988），导致抽样产生偏差。多年以来，这一不利面已得到相当大的改善。例如，到1986年为止，美国一些州的电话覆盖率已经超过了95%（Frey, 1989），而最近的平均覆盖率估计已达到95%~97%（Lavrakas, 1993）。本章将描述电话调查过程（见表11.1中所示概要），重点强调选择电话调查的标准、抽样程序、对访谈者的选择标准、培训和监督。

表11.1 实施电话调查的要点

设定调查的目的和目标
在各种数据收集方法中选择电话调查法
● 评估4个方面：管理、抽样、测量、数据质量； ● 评估电话调查是否为合适的选择。
产生抽样集合。
设计问卷和调查程序
招收、初始雇用、培训和最终选择
● 招收和选择满足初选标准的访谈者； ● 进行定向训练、技能培养、访谈者偏差训练以及反馈和评估； ● 在行为标准的基础上选择最后的合格者。如果可能，可以向其他未被选上但希望留在项目内的不合格者委以其他任务。
进行初试和项目分析，对问卷和调查程序进行修订。打印问卷和其他表格，在一个集中的环境中进行调查。
● 在工作中继续培训，每日进行简单提醒； ● 执行现场录像指导和监督； ● 根据行为标准对操作进行评估，给予及时反馈。
收集结果
报告研究发现
对结果和费用作评估、决定下一步进程等。

本章的第一部分旨在为你选择电话调查法时提供一个参考框架。鉴于电话调查的抽样方法与其他调查方式甚为不同，本章将以部分篇幅探讨在家庭（households）之间和家庭内部进行抽样的特殊技术。访谈者在电话调查的过程中扮演着极为重要的角色，这将反映在本文对访谈者的选择标准、培训和监督的强调之中。

最后，本文将简单讨论电脑在电话调查中的使用，以及电讯技术的发展给电话调查带来的挑战（或威胁）。其他重要的主题，例如问卷的制作和项目分析，已在本书第7、第10章作过探讨（分别由Goddard和Villanova及Vaux编著）。这两章陈述的原则也对电话调查有效。但电话问卷与邮件问卷稍有不同。关于设计电话调查的细节，可以参考Lucas和Adams（1977）、Cannell（1985）及Thornberry（1987）所述问卷样本。

电话调查的选用标准

在过去，调查者在不同数据收集方法之间作选择并不是太棘手的事，因为较为高级的方法仍有待进一步开发。如今面对繁多的实验证据和技术以及通讯模式的迅速改变，对一个调查方法（如电话调查）或者混合方法（如电话和邮件调查的混合）的影响，你可能被众多的选择所困扰和迷惑。实际上，方法的选择有赖于你的目标、资源、样本的特征、样本框以及不同方法的各自优势。

虽然有些研究（Groves & Kahn, 1979; Mangione, Hingson, & Barrett, 1982; Siemiatycki, 1979）已经对两个或多个调查方法（主要是电话调查、个人访谈、邮件调查）作了比较，但是这些标准很少放在具体的研究情况下进行讨论。最重要的是，在这些研究中，一些外在的因素很难控制（如样本集合、调查者的训练程度、问卷措辞等，并不能在不同方法之间保持一致），这些混淆因素以及内部效度（如不同标准的采用）和外部效度（如研究所能概括的论题和情境）问题，严重地限制了关于“单纯模式效应”（Biemer, 1988）的结论。在这一章的其余部分，在讨论重要的标准之外，我将针对电话调查的“相对”优点和缺点对以往的研究进行总结。本文将帮助决策者斟酌选择电话调查法的潜在益处。

关于电话调查的评价

通过与其他调查方法在多项标准上进行对比，我们可以对电话调查的特点有比较全面的认识。有两个可替代电话调查的方法，个人访谈和邮件调查。因为使用它们的广泛性以及技术上获得结果的有效性，我们将把它们拿来与电话调查进行比较。本节将在4个方面探讨这些标准：管理（如费用、人力、培训），抽样（如未覆盖、未应答），测量过程（如数据收集需要的时间和问卷的长度、问题类型），数据质量（如应答效度、社会期望）。

管理方面

费用。虽然在一项研究课题中需要对电话调查进行预算，但是在研究文献中很少有信息涉及一般管理费用，非直接费用和意外费用（例如，给被调查者回3~6个电话）（Frey, 1989）。通常，电话调查的费用随着样本大小的增大和对高返回率的要求而增加。为了达到这些目标，研究者必须雇用更多的人员，做更多的拨打（例如，在不同时间、不同的日子拨打某个号码以便找到被调查者）、花更多的时间准备数据库等。如果调查包括一个对被调查对象的初始筛选（如40~55岁之间的女性职业人员）；或者采用一个较长的问卷，那么更多的时间和费用将要在寻找合适的被试或进行访谈的过程中。联系居住在不同地理区域的被调查者也要更多的花费。事实上，选择专用线路（如WATS线），可以大幅度减少通话费用，因为这种线路为用户提供了拨打国内任何地方的低费用服务。如果电话限于一个或一些特定的城市或地区，你可以考虑雇用一些本地助手，以省去长途话费（虽然这会增加培训和监管的费用）。

由于每次访谈所需的人工费用较低，与个人访谈相比，电话调查操作起来就相对便宜。一般来说，电话调查比个人访谈便宜了大约 $1/3 \sim 2/3$ （Lucas & Adams, 1977; Siemiatycki, 1979）。Frey（1989）提出证据表明，电话调查比邮件调查贵 $1/6 \sim 1/2$ 。对邮件调查，即使样本较为分散或问卷较为冗长，通常也不会导致费用的重大上升，但这些因素都会引起电话调查的管理费用的增加。

员工和培训。与个人访谈相比较，在同样大小样本下，电话调查需要的员工较少。根据经验，一个典型的电话调查大约需要10位访谈者，外加2~3名监督者。实际上，在项目的开始阶段，需要征集两倍于所需的访谈者，以弥补由于技能的缺乏或兴趣的丧失而导致的人员缩减。Fowler和Mangione（1990）估计，大约要对一个访谈者进行20~30个小时的培训，才可以使他们的操作达到可接受的水平。相反，邮件调查需要较少的人员，培训监督需要的也较少。

抽样方面

未覆盖率。电话调查的一个主要缺点在于未被覆盖的家庭完全没有被抽样的机会，这可能导致结果的无效。家庭未被覆盖的原因包括：未被号码簿所收录、新增或新变更号码以及未曾申请电话服务。采用高级的随机数字拨号抽样技术（稍后再加说明），未收录的、新的、变更的号码仍旧有可能被抽样。不过，对于没有申请电话服务的家庭，电话访谈则完全不可能。一项被Lavrakas（1993）引用的1990年的官方报告表明，有10个州（如阿拉巴马、得克萨斯和西弗吉尼亚等州）家庭电话的覆盖率只有85%~90%。总体来说，居住在偏远地区和城市棚户区的居民电话覆盖率较低，其他人群，如低收入者、失业者、年轻人（如24岁以下）或者非裔美国人或墨裔美国人（Cannell, 1985b; Trewin

& Lee, 1988)也有较低的覆盖率。Hagan (1993)指出,西南地区的墨裔美国人的电话覆盖率低到65%。另外,电话未覆盖率在大家庭(如6个或以上)和单亲家庭中也比较高。不过,Thornberry和Massey (1988)曾报告,当保持家庭收入、年龄、和教育状况为常数时,种族、婚姻状况、地理区域、家庭大小以及就业状况并不对电话覆盖率构成较大的影响。如果研究者感兴趣的变量与以上这些受未覆盖率影响的特征(如健康)有密切相关时,他们的研究结果很可能会高估或低估某些问题。

未应答。不像未覆盖的家庭,未应答家庭是已被抽中作为样本,但对于部分或全部的调查问题未作应答的家庭。未能从所选择的家庭获得数据的原因可能是无能力(如被抽中的被试有生理、精神或者语言上的缺陷,以至于无法提供信息)、无法取得联系(如给所选的被试打6次电话都未能找到)或者被回绝(Groves & Lyberg, 1988)。一般而言,未应答的人倾向于年老和受教育程度低(Weaver, Holmes, & Glenn, 1975)。访谈者的态度也会影响未应答率,这方面的问题将在后面有关培训的部分加以讨论。

邮件调查中的未应答率要比电话调查高。“冷”电话调查(如打电话之前未做通知)的未应答率比个人访谈略高一些(大约高3%~5%)(Cannell, 1985b)。不过,这个差别可以通过一些预先沟通的方法(如打电话之前先通知)来消除(Fowler, 1984)。

较高的未应答率可能产生潜在的管理上的困难(如重拨、用于劝说的时间、开支)以及对调查结果效度的威胁(如统计结论效度、结构效度或外部效度)。如果未应答者与应答者在一些个体特征上系统地有所不同,调查结果可能产生偏差。因为所有的调查都是自愿参加的(或至少应该是这样),所以上述这些问题相当棘手。

必须注意的是,未应答率并不等于未应答偏差。未应答偏差的大小既取决于未应答率的高低又取决于未应答者与应答者之间真正的不同(Groves & Lyberg, 1988)。因此,只要未应答者与应答者之间并无不同,未应答率高,未必表明未应答偏差的存在。同样,只要未应答者与应答者存在着很大差异,极低的未应答率(比如说5%),在理论上也可以导致较大的未应答偏差。

电话调查中出现的另一个问题是如何计算应答率。可惜这方面没有普适的原则。Groves和Lyberg (1988)曾建议,对于不同的目的,研究者应该采用不同的方法来度量应答率,而不是用某一种受欢迎的方法。Groves和Lyberg给出了一个计算反馈率的广泛清单,包括合作率、接触率和拒绝接触率。举例来说,合作率以完成访谈的人数除以接受访谈(完成或未完成)与回绝访谈的人数总和来计算。

测量方面

数据收集和访谈的长度。电话调查的一个优点在于,访谈总耗时比较短。电话调查通常可以在一个较短的时间内完成(几天到几个星期)。但是,这种方法被指责为不能执行长达30分钟以上的调查(Lucas, & Adams, 1977)。Collins、Sykes、Wilson和

Blackshaw (1988) 也发现, 较长的访谈与较高的回绝率有关。但是他们的发现与Frankel和Sharp (1981) 的研究结果相矛盾。Frankel和Sharp (1981) 的研究指出, 访谈的长度与回绝率并无关系。Frey (1989) 也列举了几个成功的长时间 (平均50分钟) 电话调查。我怀疑电话调查的时间限制与被访谈者对话题的兴趣、被访谈者的类型 (如老人)、访谈者的经验、调查的性质以及调查的程序有关。

访谈过程。在涉及访谈的调查中, 调查者的工作是提问, 对应答者过于简短或不完整的回答进行追问, 记录应答者提供的信息。他们应该用合适的语调正确地宣读问题, 对不完善的回答进行追问时所给予的指导和反馈不应太多也不应太少, 并且记录未经主观诠释的信息。即使是一个受过良好训练的访谈者, 且按照标准的访谈程序, 其操作也会在不同的访谈中发生变化。访谈者操作的变化 (如可信度) 可因访谈者与被访谈者、监督者、同事的相互影响而产生, 也可因访谈者生理或情感的状况改变而发生, 或由访谈者自身的性格气质使然。尽管实际上不可能完全消除访谈过程中的这些问题, 较为严格的监督、迅速的反馈、每日的报告和现场指导都可以减少这些问题。

匿名/保密和问题类型。在大多数情况下, 被访谈者的姓名、电话号码、甚至地址对于电话调查和个人访谈都是已知的。因此, 这些调查方式实际上无法在匿名情况下进行 (大多数邮件调查并不存在这个问题)。既然被访谈者的身份已知, 他们对敏感的、使人感到有危险性或者不安、为难的问题的回答难免就会在调查中有所歪曲 (Lucas & Adams, 1977)。甚至当调查者向被访谈者做出保证严守秘密的情况下, 所收集到的数据的质量也可能因为被试的怀疑和担心而受影响 (Frey, 1989)。

就问题的类型来说, 对于复杂的问题, 使用个人访谈较之电话调查更为合适。在邮件调查中也不宜采用复杂的问题。Cannell (1985a) 建议, 研究者应该尽量把问题做得简单, 因为一般来说, 应答者在电话里对问题的理解会有些困难。使用较为简单的问题也可以限制因追问不足而导致的访谈者偏差。就像在下文将要讨论的那样, 通常要花大量的时间来培训访谈者, 以便他们能对开放式的问题进行令人满意的追问。这是因为: (a) 访谈者要有发现不充分的回答的能力; (b) 对开放式问题中频繁出现的缩短回答, 要进行更多的追问 (Frey, 1989)。但是, 开放式问题的用途和效果亦不应被低估或抹杀。被访谈者对开放式问题的回答经常给所研究的问题带来新的启示或者催生新的研究想法。

数据质量方面

评价数据质量的标准包括精确度 (或应答效度)、社会期许的避免、未应答项目、开放式问题所得到的信息量以及采用不同方法所得到的应答的相似程度。De Leeuw 和 van der Zouwen (1988) 对以前的研究做了元分析并得出结论: 个人访谈与电话调查在上述标准上差异很小, 不过个人访谈要略强些。事实上, 他们发现, 这两种方法得到的数据

质量方面的差异正在变得越来越小,这种趋势可能是因为电话调查经验的增加造成的。其他的数据质量指标还包括应答风格(如默许、回避、走极端)和心理测量属性(psychometric properties)(如测试一再测试或者内在一致性信度估计)。例如,Jordan、Marcus和Reeder(1980)发现,电话调查比个人访谈更多地出现赞同、极端或者回避式的应答。不过,这两种调查方法的心理测量属性并没有不同(Aneshensel, Frerichs, Clark, & Yokopenic, 1982; Herman, 1977)。

选择电话调查的决策帮助

如果一项研究计划只允许选用单独一个调查方法,研究者应该对每种方法的优缺点进行评估。在大多数情况下,没有绝对正确的选择。下面提供的决策帮助仅供参考。考虑方法的时候,首先应该考虑在邮件调查和与访谈有关的调查(如电话调查、个人访谈或者两者的结合)之间进行选择,看看哪一类方法更有利于达到研究目标。这个决定可以通过对以下问题的回答来作出:

1. 是不是许多问题需要根据应答者对先前问题的回答而临时决定?
2. 对问题的应答必须控制在一个特定的时间内进行吗?
3. 调查的问题很复杂吗?
4. 是否可以得到全面精确的邮件通讯录和具有代表性的参与者?
5. 调查是否非常敏感、具有令人感到威胁或为难之处或者对个人隐私涉及较多?
6. 是否样本分布在不同的地理区域?

如果调查中提出的问题依被调查者对前一个问题的应答而定,则应考虑采用访谈式的调查。同样,当调查问题比较复杂、全面、精确的邮件列表不可得或者收集数据要求在某一天的特定时间进行(如早上)时,采用访谈式的调查方法也是比较合适的。除了一些特殊情况(例如,当采用经验抽样法时用BP机或定时器提醒应答者时),否则安排在特定时间进行邮件调查是相当困难的,因为研究者无法控制调查问题何时被答。

事实上,如果调查问题非常敏感或者应答者难以接近(例如,分别住在加利福尼亚、佛罗里达、夏威夷、阿拉斯加或加拿大和日本关岛),访谈式的调查方法并不是最好的选择。由于在访谈中完全的匿名几乎不可能,被访谈者倾向于尽早结束访谈或者避免回答敏感问题。重申研究的保密性可能减少回绝者,但这样得到的数据也会出现质量问题。当样本分布广泛的时候,花在长途电话上(甚至使用WATS专线)或者探访上的开支都会大大增加。

如果认定访谈式的调查是更好的选择,接着是决定哪一种访谈(电话或个人访谈)比较合适。

下面的问题有助于做这个决定:

- 1.数据是否要在特殊的地方收集？比如工厂、超市或邮电所？
- 2.访谈者需要向应答者展示一些东西吗？
- 3.是否根据行为或者生理特征来选择合适的应答者？
- 4.访谈中是否采用许多问题？
- 5.是否某种类型的人被排除在外？
- 6.数据是否必须在非常短的时期内收集完成？
- 7.应答者是否居住在广泛的地理区域内？
- 8.应答者周围的人在访谈过程中是否可能对其发生影响？

如果应答者必须在一个特定的地方被访谈或者必须看到出示的某些物品（如产品包装或者汽车的颜色），电话调查并不是合适的选择。同样，如果对应答者的选择是根据他们的生理特征（如金头发），或者行为（如嚼烟草），或者如果调查采用一份较长的问卷（例如，要花60分钟以上的时间来完成），也不宜采用电话调查。某些地理区域（如西南地区的墨裔）或者某类人（如无家可归者）的电话往往不可得，如果这些人代表样本中的重要成分，电话调查就至少不应该作为一项研究惟一使用的方法。

电话调查往往是最为快捷的方法。例如，若是你的单位紧急要求你在10日内取得雇员们对于一个泊车位新分配系统的意见，你可能会花4天时间通过电话完成300个访谈。你若用其他访谈方式就可能很难完成任务。电话调查也适用于应答者居住分散的情况，或者应答者周围的人可能影响他/她应答的场合——因为电话调查中应答者周围的人不会听到所问的问题，从而在一定程度上降低了应答者保护面子或自我表现的需要。

产生样本组

在进行调查前，研究者需要决定该对多少人、什么样的人进行抽样。不应首先考虑所要调查的样本，而是先要着眼于确定所研究的总体（样本取之其中）。换句话说，研究的目的是通过抽取的样本来了解总体。从样本到总体的结论推广，在很大程度上取决于抽样计划（例如，抽样框和样本大小）。在本书的其他部分（McCready，第8章）可以得到关于抽样程序的大致信息和相关资料。

电话调查包括两个完全不同但不可分割的抽样计划：在家庭之间抽样和在家庭之内抽样。前一个是用电话号码簿或其他途径抽取有代表性的家庭（例如，电话号码所代表的家庭），后者是在已被选择的家庭内选择应答者。

在家庭之间抽样

在家庭之间抽样是研究者通过电话簿、随机数字拨号（RDD）或者Waksberg两步

RDD抽样等方法随机产生电话号码的一种过程。通讯录中的号码通常可由电话公司、组织机构（如雇员花名册）、协会（如美国心理学协会会员录）或者商业性的目录提供者等处获得。现在也可以买到一个CD-ROM的数据库（花不了50美元），其中包含了7 000万个家庭。事实上，对于公众来说，电话号码簿以外的通讯录是不易得到的。并且，电话号码簿往往包含重复的或者不正确的电话号码、不能接通的号码以及未能收入可用的应答者（新号码的、未收入号码的、才变更号码的或者未装电话的可用应答者）。未收入可用应答者将对统计结论和外部效度构成威胁。

可以用电脑或者随机数表（random-numbers table）进行随机数选择，从一个或多个目录中产生样本组。样本组的大小依赖于你的资源以及你所容许的抽样错误的大小。对于分层抽样或区域概率抽样，电话号码可以根据地理区域或者其他特征（如公司大小、性别、工作性质）确定。抽样组也可以通过加一位（Add-a-Digit）抽样法进行，这是一种通讯录辅助技术。这个方法要求研究者在电话号码的最后4位数码上加上1个（或2个或3个）常数（或随机数）。如果数码593-0822被抽取，当一个数字9被随机指定后，实际拨打的号码将是593-0831。一般而言，通讯录抽样会由于目录的不完全而导致取样偏差，甚至加一位的取样方法也不能将其完全消除。

为了减少因为号码簿不完备而导致的偏差，研究者逐渐发展了RDD抽样以及它的几个修订本（cooper, 1964; waksberg, 1978）。RDD法无需在抽样之前列出家庭单位（Frey, 1989）。任何已给出的电话号码，比如614-593-0922，皆由三部分组成：区号（如区号614）、前缀（prefix）或叫中央办公室码（如前缀593）以及后缀（0822）。统共有640个3位数前缀分配给每一个区号，统共有1万个后缀（从0000~9999）分配给每一个前缀，实际上，每个前缀并没有连接所有可能的后缀。后缀通常是以1 000或100为单位成捆地进行分配。例如，一捆1 000个后缀号码可能分配给一所大学，或者一捆100个后缀号码可能分配给一家宾馆。人们可以从电话号码簿或者地方电话公司的信息上发现这一点。

进行RDD抽样时，首先决定样本被抽取的地区区号，然后任意选取前缀，对每个前缀内的后缀进行任意抽取。任意数字的产生可以通过随机数表或者电脑来进行。Lavrakas（1993, pp.44-46）提供了一套用BASIC语言写出的可以随机产生后缀的程序。Lavrakas也探讨了与RDD抽样有关的其他重要信息，如确定未被使用的后缀区以及为每个前缀确定用户使用号码。

Waksberg（1978）提出了一个已经在大规模调查中频繁使用的两步RDD抽样法。第一步，研究者随机选择少量区号与前缀的组合（大约50~100）。然后，在每个区号—前缀的组合之后添加两位随机数字。由这一系列由8位数组成的号码组成初级抽样单位（PSU_s），以最终产生抽样组。每个初级抽样单位（PSU），可以产生100个号码。例如，可以从初级抽样单位（PSU）614-593-08得到由614-593-08-00到614-593-08-99的100

个电话号码。在两个随机尾数（如22）加入PSU之后，最终的拨打号码变成614-593-08-22。如果这个数字是一个居民的电话号码，PSU将被保留下来进行第二步操作。否则，PSU被放弃。在第二步，随机产生的两位数加到PSU之后，直到获得一定数目的居民电话号码。总体而言，这个抽样技术是有效的和经济的，但以增加人员的初始工作量和取样精度的少量损失为代价（Frey, 1989; Groves, 1989）。

在家庭内抽样

当一个访谈者拨打产生的电话号码时，他或她可能立即面临一个困境：在这个家庭内谁应该被访谈？关于如何决定应答者，可以通过下面步骤进行。最简单的方法莫过于径直访谈接电话的人——如果他/她符合事先定下的标准（例如，21岁以上的男人）。事实上，这种方法降低了男性的代表率，因为女性倾向于先接电话（Frey, 1989）。最初以个人访谈为目的，Kish（1949）采用了一个方法，要求接电话的人罗列所有家庭成员的姓名、性别、年龄，然后访谈者给每个成员根据年龄和性别分配一个数码，候选人就是根据这个数字表来选择。Kish的方法及其修改版的介绍详见Lavrakas（1993）和Frey（1989）以及Kish（1949）。Kish的步骤可能在家庭内消除了未覆盖误差。既然在有些家庭并非所有可能应答者皆被列举（例如，应答者的拒绝或者漏报、调查者的疏忽），少量的未覆盖错误仍然可能发生（例如，最年青的家庭成员很少被抽取）。尽管减少了未覆盖错误，这个方法因抽样过程存在大量的强迫性的问题而产生较高的非应答率。Frey总结了Kish的方法，认为它较之于电话访谈，其实更适用于个人访谈。

另一个在家庭内抽取应答者的技术被称作最近（或下一个）生日法，这个方法也是一种概率抽样技术，因为人的生日是随机的。进一步说，因为每个家庭总有一个人刚过或者即将过生日，性别和年龄偏差就被消除了。因其对被调查者只有较少的介入或强迫性以及使用的方便，这种方法自20世纪80年代中叶以来在电话调查中获得广泛的应用。

人 员

若是对访谈者没有足够的训练和现场工作监督，电话调查是不能顺利进行的。在讨论对访谈者的培训和监督之前，两个经常被忽略的但又是关键的问题需要先讨论一下：即对访谈者的招收和初选。

对访谈者的招收和初选

如果有1000个访谈，通常需要10个访谈者，尽管比率可能会因预算、样本大小、时间要求、可获得的工作空间以及调查长度而有所不同。由于潜在流失的存在，应该先招收两倍于终选数目的访谈者。建议向申请者介绍将来工作的实际情况。介绍的内容应该

涉及结果和奖励（如报酬、奖励、评价、推荐信、著述权）、工作小时和工作负担（如每晚两小时，8~10时）、将被访谈的人数、选择标准（如工作经验、表达的清晰度、对培训的评估）、责任（如监督办法或职业准则）、调查范围以及其他要求（如与培训有关的规则、迟到、未到、抽烟、食物、饮料等）。对将来工作正反两面的介绍，可能促使申请者重新评估他们的决定，另外，也促使他们在加入培训计划之前对不现实的期望予以调整。申请者因这些介绍而产生的自我选择使你的筛选更为容易。

如果潜在的访谈者决定申请这份工作，通过模拟任务或者角色扮演对申请者的阅读和写作能力进行估计将有所裨益。人口统计特征（如种族、性别、年龄、宗教），除非有证据表明它们对访谈的结果有所影响，不应该被用作选择访谈者的标准（Lavrakas, 1993）。Fowler和Mangione（1990）提出，访谈者应该具备较好的读写能力及乐观理性的性格。他们进一步指出：似乎没有其他“区分潜在访谈者的有效选择标准（p.140）。”但是，其他的因素也可能存在。例如，Groves和Fultz（1985）发现，过去的工作经验是减少回绝率的重要因素。既然很少有证据表明何种特殊技能（如说服人的技能、交流技能）、能力（如表达的清晰度、语调），或者个人特征（如友好、对拒绝的承受力）对于访谈工作是至关重要的，Fowler和Mangione的关于选择标准的结论应该慎加运用。

培训与终选

初选之后，应该是一系列训练阶段（大约20~30个小时），这些训练不应考虑潜在访谈者的以往经验，以使得访谈程序标准化。训练的内容应该包括怎样与被访谈者接触并得到他们的合作、如何与被访谈者建立良好的关系、如何保证访谈结构完整、如何在避免个人观点介入的情况下记录回答。培训可能需要使用访谈手册、演讲/讨论、示范、交互式的计算机辅助程序、角色扮演、模仿和有监督的练习。对访谈者进行培训指导的范例参见Frey（1989）。

培训有四重目的：指导被培训者了解调查范围、调查问卷和调查程序；使被培训者提高访谈技能，练习合适的言谈方式以及学会如何进行追问和获得反馈；减少潜在的与访谈者有关的偏差；评估被培训者的操作，并以此作为选择最后雇用人员的依据。以下的部分将描述如何通过培训达到这些目的。

适应性培训

在适应性培训期间，被培训者应该知道调查的目的，了解将要访谈的应答者的类型，知道一个出色的访谈者所应具备的特点及所应遵守的职业道德，知道电话号码是如何产生和使用的以及了解其他的管理规则。另外，他们应该学习何时与怎样正确使用调查联系记录单、介绍性陈述、备用陈述以及回绝报告表（样表参见Lavrakas, 1993）。应该给被培训者足够的时间，以讨论和解释问卷采用的项目。访谈者在访谈中不明白自己

该问的问题，从而导致困惑和错误的情况并不少见。

技能培训

当访谈者进行电话调查时，每天会碰到很多回绝，被培训者应该学会怎样谋求应答者的合作且与他们在整个访谈期间保持良好的关系。他们还应该了解回绝的原因，知道怎样去劝说受访者又不让其感到有所为难，知道如何处理他们自己对回绝的畏惧。关于说服技巧的信息参见Groves（1989）。

除了上面指出的技巧，被培训者应学会改进声音质量。访谈者有时变得漫不经心，以致未能使用合适有效的发音（Strafford & Grant, 1993）。Okenberg和Cannell（1988）发现，当访谈者在访谈中表现得有能力和信心、言语标准、发出的语调低沉且声音大以及说话的速度较快，都能够减小回绝率。Strafford和Grant进一步建议，访谈者应该在声音中保持“微笑”（尽管受访者并不能看到），以便使交流得到润滑。

在电话访谈中，因为没有视觉线索，访谈者经常试图从受访者那里获得反馈以确信交流的效果。这种寻求反馈的愿望可以从访谈者比平时说得多或者说些不该说的话等现象反映出来（Connell, 1985b）。在培训期间，受训者应该知道，一些简单的句子如“我理解”或者“我明白了”，已经足以保持交流。“我明白了”这类句子可以打印在问卷的每个问题之后，让访谈者在受访者答完每个问题之后念出来。

访谈者偏差培训

电话调查的结果是通过访谈者向应答者读出问题、对模糊的或不完全的应答进行追问，以及记录下最后的回答这些方式获得。歪曲的结果是很容易出现的，甚至在问卷设计得相当完善的情况下也是如此，因为一个或多个访谈者皆可能出现偏差。举例来说，访谈者可能没有精确地读出问卷所述问题或未能在必要的时候对回答进行追问。访谈者对受访者的追问过于直露（它可能影响应答的清晰度或修饰程度）或用自己的判断来记录应答，这些情况下也会发生误差。其他的偏差包括人际行为的不恰当（如显得有侵略性）、对应答者作出评价性的反馈（例如，“这很好”或“你很幸运”）。这些潜在的错误可以通过访谈者偏差培训而减少。事实上，有证据表明，没有以往经验的受训者需要大量的时间去掌握追问技巧，对于采用开放式问题的问卷更是如此。Fowler和Mangione（1990）报告，在他们的研究中，在历时10天的培训之后，只有47%的被培训者显示了极佳的或令人满意的对开放式问题进行追问的能力。

反馈和评估

整个培训期间，应该给受训者较多的机会去通过个案研究、行为模仿、角色扮演、讨论/批评等方法来练习上述的技能。培训之后，被培训者应该在有指导的情况下做3~5

个访谈。在他们不知道的情况下，让他们访谈同一个被试。每一个“同谋”受访者应该根据事先设计好的范本回答问题，然后就一系列关键行为对被培训者的工作进行评估。这些行为标准包括——但不限于——言语清晰度（项目之间的时间间隔、发音错误、不恰当的词尾变化或不恰当的语调）、被读错的问题的数目、不必要或不正确重复问题的数目、过于直露追问的的数目、不正确跳题的频度、忘记对不清晰问题进行追问的频度，对问题作不必要解释的频度、对所答问题记录模糊的频度、予以不恰当举例或反馈的数目、不适当的人际行为的数目、不恰当情绪表露（大笑）的数目、对调查联系清单或其他表格上的要求有所违反的数目（Cannell, 1985a; Fowler & Mangione, 1990）。应该强调的是，这些标准应被可靠地测量。这些标准不仅使访谈者的选择有较为客观的标准，也使整个培训的效果（例如访谈者工作的可靠度）可以被评估。进一步说，这些标准和其他标准（如同一个访谈者的未应答率、处理记录的速度）可以用来在正式调查开始后指导和评估访谈者的工作。在一个学术机构中，不合适的受训者仍可以委以其他工作（如打印，数据输入）——假如他们希望留在课题组内的话。

总 结

由于高级电讯技术和计算机技术的发展，对电话调查的新挑战不断出现。例如，电话调查的使用者正在面临留言机、来电显示器、按键式数据输入（TDE）或声音识别器（VRE）、来电转接、汽车电话、手机和电视电话的挑战。这些新产品非常可能带来诸多问题，如非应答率的增加和抽样偏差。事实上，这些新产品的潜在冲击至今仍不是很清楚。

另一个挑战来自过去25年中计算机辅助电话访谈（CARI）程序的革新。Nicholls（1988）总结了CATI的3个主要功能：辅助电话调查（如减少开销），提高和控制调查数据的质量（如减少错误），能实现变化性强的调查设计（如问题的随机化）。在一个典型的CATI程序中，访谈者采用由显示屏中出现的问题，在应答者答完问题之后将所有信息立即键入电脑。有些CATI程序对调查问题的显现次序和表达用词进行随机处理或者根据前面的应答输入选择下一个问题，这可以使一部分访谈偏差得以消除。

有些CATI程序能够对应答的自相矛盾之处进行跟踪，能够随机产生电话号码，可以安排打电话和回电事宜，或者对访谈进行记录。毫无疑问，只要使用者训练有素，程序运行正常，CATI程序就能为研究者的工作提供巨大的方便和灵活性。但是，对于CATI的许多优点仍存在争论，有待经验研究的支持（Porst, Schneid, & van Brouwershaven, 1994）。关于使用CATI后的数据质量、非应答率、访谈偏差、软件评估以及培训等的研究有待进行。应该强调，CATI对由电脑焦虑、问卷设计不当、抽样程序的不合适、访谈过程的不标准、对受访者的选择性偏差，访谈培训和监督的不足、访谈者与应答者的关

系不善等因素而产生的问题是无能为力的。

推 荐 阅 读

最后,应当指出的是,并没有哪一种数据收集方式能够回答社会科学所关心的所有问题。每一种方法都有其局限性。但是,通过使用多种方法(如三角测量法),研究者将能够从不同的角度考察问题。

这里推荐4本书,以帮助读者对访谈式研究作进一步的全面了解。关于访谈式调查,参见Fowler和Mangione著《标准化调查访谈》(*Standardized Survey Interviewing*, 1990)和Groves所著《调查偏差和调查费用》(*Survey Errors and Survey Costs*, 1989);关于“使用电话的访谈式调查”参见Frey著《采用电话的调查研究》(*Survey Research by Telephone*, 1989)和Lavrakas著《电话调查法》(*Telephone Survey Methods*, 1993)。这些书中探讨了一些重要的问题,例如抽样、标准化、问卷设计、说服技巧、培训以及CATI程序的使用等。

参 考 文 献

- Aneshensel, C. S., Frerichs, R. R., Clark, V. A., & Yokopenic, P. A. (1982). Measuring depression in the community: A comparison of telephone and personal interviews. *Public Opinion Quarterly*, 46, 110-121.
- Biemer, P. P. (1988). Measuring data quality. In R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls II, & J. Waksberg (Eds.), *Telephone survey methodology* (pp. 273-282). New York: John Wiley.
- Cannell, C. F. (1985a). Experiments in the improvement of response accuracy. In T. W. Beed & R. J. Stimson (Eds.), *Survey interviewing: Theory and techniques* (pp. 24-62). Boston: George Allen & Unwin.
- Cannell, C. F. (1985b). Interviewing in telephone surveys. In T. W. Beed & R. J. Stimson (Eds.), *Survey interviewing: Theory and techniques* (pp. 63-84). Boston: George Allen & Unwin.
- Collins, M., Sykes, W., Wilson, P., & Blackshaw, N. (1988). Nonresponse: The UK experience. In R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls II, & J. Waksberg (Eds.), *Telephone survey methodology* (pp. 213-231). New York: John Wiley.
- Cooper, S. L. (1964). Random sampling by telephone: An improved method. *Journal of Marketing Research*, 1, 45-48.
- de Leeuw, E. D., & van der Zouwen, J. (1988). Data quality in telephone and face to face surveys: A comparative meta-analysis. In R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls II, & J. Waksberg (Eds.), *Telephone survey methodology* (pp. 283-299). New York: John Wiley.
- Fowler, F. J., Jr. (1984). *Survey research methods*. Beverly Hills, CA: Sage.
- Fowler, F. J., Jr., & Mangione, T. W. (1990). *Standardized survey interviewing: Minimizing interviewer-related error*. Newbury Park, CA: Sage.
- Frankel, J., & Sharp, L. (1981, January). Measurement of respondent burden. *Statistical Reporter*, pp. 105-111.
- Frey, J. H. (1989). *Survey research by telephone* (2nd ed.). Newbury Park, CA: Sage.
- Groves, R. M. (1989). *Survey errors and survey costs*. New York: John Wiley.
- Groves, R. M., & Fultz, N. H. (1985). Gender effects among telephone interviewers in a survey of economic attitudes. *Sociological Methods and Research*, 14, 31-52.
- Groves, R. M., & Kahn, R. L. (1979). *Surveys by telephone: A national comparison with personal interviews*. New York: Academic Press.
- Groves, R. M., & Lyberg, L. E. (1988). An overview of nonresponse issues in telephone surveys. In R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls II, & J. Waksberg (Eds.), *Telephone survey methodology* (pp. 191-211). New York: John Wiley.
- Hagan, F. E. (1993). *Research methods in criminal justice and criminology* (3rd ed.). New York: Macmillan.

- Herman, J. B. (1977). Mixed mode data collection: Telephone and personal interviewing. *Journal of Applied Psychology*, 62, 399-404.
- Jordan, W. H., Marcus, A. C., & Reeder, L. G. (1980). Response styles in telephone and household interviewing: A field experiment. *Public Opinion Quarterly*, 44, 210-222.
- Kish, L. (1949). A procedure for objective respondent selection within the household. *American Statistical Association Journal*, 44, 380-387.
- Lavrakas, P. J. (1993). *Telephone survey methods: Sampling, selection, and supervision* (2nd ed.). Newbury Park, CA: Sage.
- Lucas, W. A., & Adams, W. C. (1977). *An assessment of telephone survey methods*. Santa Monica, CA: The Rand Corporation.
- Mangione, T. W., Hingson, R., & Barrett, J. (1982). Collecting sensitive data: A comparison of three survey strategies. *Sociological Methods and Research*, 10, 337-346.
- Nicholls, W. L., II (1988). Computer-assisted telephone interviewing: A general introduction. In R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls II, & J. Waksberg (Eds.), *Telephone survey methodology* (pp. 377-385). New York: John Wiley.
- Oksenberg, L., & Cannell, C. (1988). Effects of interviewer vocal characteristics on nonresponse. In R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls II, & J. Waksberg (Eds.), *Telephone survey methodology* (pp. 257-269). New York: John Wiley.
- Porst, R., Schneid, M., & van Brouwershaven, J. W. (1994). Computer-assisted interviewing in social and market research. In I. Borg & P. Mohler (Eds.), *New trends in empirical social research* (pp. 79-98). Berlin: de Gruyter.
- Siemiatycki, J. (1979). A comparison of mail, telephone and home interview strategies for household health surveys. *American Journal of Public Health*, 69, 238-245.
- Strafford, J., & Grant, C. (1993). *Effective sales management* (2nd ed.). London: Butterworth-Heinemann.
- Thornberry, O. T., Jr. (1987). *An experimental comparison of telephone and personal health interview surveys*. Washington, DC: Government Printing Office.
- Thornberry, O. T., Jr., & Massey, J. T. (1988). Trends in United States telephone coverage across time and subgroups. In R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls II, & J. Waksberg (Eds.), *Telephone survey methodology* (pp. 25-49). New York: John Wiley.
- Trewin, D., & Lee, G. (1988). International comparisons of telephone coverage. In R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls II, & J. Waksberg (Eds.), *Telephone survey methodology* (pp. 9-24). New York: John Wiley.
- Waksberg, J. (1978). Sampling methods for random digit dialing. *Journal of the American Statistical Association*, 73, 40-46.
- Weaver, C. N., Holmes, S. L., & Glenn, N. D. (1975). Some characteristics of inaccessible respondents in a telephone survey. *Journal of Applied Psychology*, 60, 260-262.



第 12 章

从团体中收集数据

Stephen J. Zaccaro

Michelle Marks

人类的行为总是发生于某个特定社会背景之中，而行为常常受到其他个体存在的影响。一方面，这种影响可能是相当被动的，仅仅是他人的存在便可影响行为的频率、强度和适当性（Zajonc, 1965）；另一方面，社会情景对行为的影响也可能是整体性的，一个个体的反应完全依赖于其他个体的反应（Orasanu & Salas, 1993）。在上述两种情况下，或在任何其他一种社会互动模式下，如果不考虑社会互动的作用，个体的行为就不能被完全理解。因此，进行群体水平上的研究就变得极其必要。

个体和团体研究的差异主要表现在团体内个人互动的过程中，因此，进行细致的团体研究就要求考虑这些互动。这种考虑不仅表现在对团体现象的定义上，也表现在实验设计、将团体作为一个整体、从个体身上收集数据以及数据分析上。本章中，我们将提供一些团体研究的规范，主要包括：研究范围的选择以及相关的研究设计；数据收集步骤，包括团体的获得和选取以及对群体活动的测量与量化；团体水平上的数据分析。

个体研究和团体研究的根本不同之处，很大一部分在于，后者经常考察在合作与互动中工作的一些个体。集体和人际活动都可能对个体和集体的随后反应构成影响。举例来说，这种影响在个体决策和团体决策之间的差异上就有所表现：个体决策包括对可能的决策情景的认知分析过程、信息组织、确定候选方案以及选择“最优”方案（Lipshitz, 1993；Montgomery & Svenson, 1989；Pennigton & Hastie, 1986）。选择过程体现了多种决策启发性思维的运用，因此可能导致多种偏差（Cohen, 1993；Tversky & Kahneman, 1972, 1974）。可以看出，在个体决策中，主要的因素是个体的认知动力。

团体决策也把上述认知动力学作为其关键前提（Duffy, 1993）。但是，典型的团体决策研究主要考察在人际互动背景下的团体成员的认知过程（Gualtieri, Parker, & Zaccaro, 1995）。例如，关于“团体思维”的研究发现，团体对最重要事务的强调会导致某种“团体一致性”，从而导致有缺陷的决策（Janis, 1982）。同时，对团体“极

化”现象的研究发现，团体决策也会导致一边倒的决策（更冒险或者更谨慎），且决策结果比个体独立思考更为偏激（Myers & Lamm, 1976; Wallach, Kogan, & Bem, 1962）。对于团体思维和团体极化的理论解释集中在有关人际互动如何影响个体，并使其在团体情景下表现出不同行为方式。从本章的讨论可以看到，团体行为与个体行为的差异使得团体研究设计与个体研究有很大的不同，这些差异包括研究变量的性质以及数据的统计分析方法。

人际互动过程和集体反应模式的存在导致群体研究的另一个特征，这个特征在分析数据时变得尤其重要。当在一个团体中成员之间存在互动（包括非言语方式的互动），这些互动的内容和性质可能显著地影响团体成员的个体反应。Kenny和La Voie（1985）称此为“非独立性”，即在一个团体中，个体成员的数据相对于其他成员的数据并不独立。由于多数统计分析都要求数据之间具有独立性（如方差分析），这就需要在处理团体数据时使用不同的统计方法（Anderson & Ager, 1978; Kenny & La Voie, 1985）。

个体水平的研究只需要单个层次的数据分析，即个体的数据。但是，当数据是从团体成员中收集时（例如，使用“多层嵌套设计”，Kenny & La Voie, 1985），就有可能对个体和团体两个层次进行了分析。一个在团体研究中常见的错误就是，在嵌套设计时，忽视了其中一个层次上的数据分析。许多团体行为可能是必须在两个水平上分析才能够得以更好的理解，即考虑集合间的差异（使用团体平均数进行团体水平分析）和集合内的差异（使用个体测量值进行个体水平上的分析）。这表明，当一个人形成假设时，就应该考虑到个体和团体这两个水平上的变异；同时，对从嵌套在团体中的个体身上收集数据时，统计分析应考虑多个水平的影响源。

上述分析表达了团体水平的研究和个体研究的另一个不同。对个体行为的研究经常把个体行为作为变异的来源。但是，在团体研究中，变异不仅来源于团体成员个体特征上，还表现在团体成员的互动上（例如，团体凝聚力、习俗压力和角色定义）。其他一些团体研究的变异则可能来自团体的结构（团体的大小、层次结构的组织、沟通过程）。团体研究的设计需要考虑到这些变异，即使这些变异可能并非是研究者所关心的变量。例如，团体大小能够影响团体凝聚力（Cartwright, 1968），团体成员从众和服从的强度（Asch, 1951, 1955; Milgram, 1963, 1965），以及团体成员为该集体付出努力的程度（Latane, Williams, & Harkins, 1979）。这样，在团体设计中，研究者必须考虑到团体大小这一因素，即便它并不是研究所关注的焦点。

选择团体研究的环境和设计

团体研究的环境在文献中得到了广泛的关注和讨论。多数研究者使用实验室环境进行实验研究。使用这种方式，研究者可以控制一些变量，能够得到一定程度上的因果关

系；而且，研究者可以创造一些在自然环境下不常出现的情景和状态（如紧张、危机）。这个策略有助于调查那些罕有的、却意义重大的事件，这些事件往往对其后该团体的规范、文化及长期集体行为构成决定性的影响。然而，研究者对这种特设团体的频繁使用往往被批评为过于人工化，缺乏外在效度（Forsyth, 1990）。事实上，一些团体现象只在团体成员有了大量的共同经验后才会产生（如团体规范、亲和力）。既然大部分实验研究团体的聚合时间甚短，也不存在对继续交往的期望，用这种团体对这类现象进行研究就很容易遭到质疑。

在团体未受干扰的自然环境下进行研究，是一个较好的团体研究策略。其中最常用的方法有两种：一种是使用问卷调查方法进行相关研究；另一种是在一段时间中观察最显著的事件。与实验室研究和特殊团体研究相比，对“真实”团体进行研究明显具有更大的普遍性和适用性。但是，这样的研究策略也存在着一些问题。相关研究并不能明确指示关键变量之间是否存在因果关系，这一缺点很大程度上制约了这种研究方法检验理论假设的能力；而且，如果调查是在某个时点上施加于团体，那么，对团体规范的出现和凝聚力的形成就不能进行测量；另外，如果数据只是从团体成员处得到的，那么，相关分析的结果可能仅仅是由统计的偏差或者反应的偏差所致。这些缺点可以通过如下方案予以克服：（1）在团体形成开始以及几个不同的关键时期，施以多个问卷调查；（2）调查数据与其他来源的数据（例如，观察和档案分析得到的数据）相结合。

观察法提供了一种对团体活动的实时测评。但是，这些方法在两个重要方面遭到了批评。首先，有限的观察时间可能无法观测到重大的团体事件或团体活动，无法充分地估计目标变量之间的假设关系。例如，如果研究者对不同类型压力下团体的反应感兴趣，往往一次极长时间的观察才可以保证研究者能够充分地分析压力源以及相应的团体反应，并充分理解这个现象，但这么长的时间来做研究是不现实的。此外，这类观察方法可能存在“霍桑效应”（Roethlisberger & Dickson, 1939），即团体成员在意识到他们被观察时，会改变自己的行为内容及频率。不告知被观察者进行观察是消除此效应的一个方法，但这样做又会面对明显的职业道德问题。

我们以及其他团体研究者建议采用如下的方法，来处理实验室/实验研究和自然/相关研究的两难困境。一个建议是，方法上的考虑必须服从于研究目的。Driskell和Salas（1992）指出，实验研究提供了在一个控制的环境中检验某个理论或假设的方法。他们认为，是实验研究的理论，而不是某个发现或者情景能在现实世界中予以应用。这样，如果团体研究的目的在于检验某个团体行为的理论，那么，实验研究就应该得到优先考虑。但如果团体研究的目的在于理解现实世界中的某个情景，或把某个理论应用于现实环境之中，那么，自然环境下的研究就应该得到优先考虑（Driskell & Salas, 1992）。

特殊目的召集的团体不必被局限于某个人为的环境中；同样，真实的团体也可以适用实验操作。研究者可以为研究目的制造出一个人造团体，然后在自然环境中进行操

纵。同样,研究者也可以给一个已存在的团体设置必要的实验情景。这样的研究实例包括在课堂环境中建立项目小组,使已存在的团队接受培训,以及在不同的组织团体中使用不同的工作条件。这些策略都可以缓解一些单纯的实验室/实验研究或者自然/相关研究所带来的问题。

一个针对团体现象的研究可能有多重目的。因此,研究者应该考虑进行一系列研究。比如,进行一些实验研究来检验理论,进行另一些在自然环境下的相关研究来考察实验研究结果的普遍性和实用性。这样一种多策略研究方法可以补偿单个研究方法的不足(McGrath, 1984),并且提供一种评价团体行为的理论及其在现实世界中适用性的基础(Driskell & Salas, 1992)。

团体研究有两种基本的数据收集方法:一种是团体水平上的数据收集。这样的设计可适用于团体大小或结构与团体合作之间关系的研究等。这类研究的因变量通常是指团队成员之间的互动产生的结果。让团体共同解决一个问题的团体决策研究就是团体水平研究的例子,在这类研究中,统计分析是以团体整体为单元进行的。另外一种研究设计是从一个团体中的个体身上收集数据,这种设计称为“嵌套设计”(Kenny & La Voie, 1985)。正如Kenny和La Voie(1985)在研究中展示的那样,统计分析可以同时从团体水平和个体水平上进行。研究者通常犯的一个错误就是,在本来更应该进行团体水平分析时,却仅进行了个体水平的分析。正如本章后面将讨论的,即使数据是从团体成员水平上收集的,且研究假设只关注个体水平的关系模式,对团体水平效应是否存在进行统计分析仍然是必须的。

数据收集程序

本书的其他章节讨论了在个体水平上收集数据的问题。但是,一些特别的问题将会出现在从团体中收集数据的过程之中。这些问题与被试的召集和变量的特定化有关。

被试的召集

在被试召集方面,实验室/实验研究和自然/相关研究是不同的。在实验研究中,研究者首先需要召集被试。在很多大学中,被试团体往往是那些入门性的心理学课程学生中的志愿者。但对团体研究者来说,重要的一点就是那些同意在一段时间内参加实验的个体的数目。如果人数不足,实验便无法完成。对此,研究者通常使用的方法就是召集多于需求数目的被试。如果参加的个体数目超过了实验所需要的,研究者便随机选择一些,剩下的被试则被要求以后来参加实验,或是参加一个与此次实验无关的研究。

如果被试在实验前就互相认识,一些实验处理将会出现混淆。在这种情况下,研究者需要在召集被试或分配被试时避免这种组合的出现。例如,在一项探究以任务为基础

和社交为基础的团体凝聚力的研究中, Zaccaro和McCoy (1988) 对被试签名表格进行甄选, 使被试的朋友不会出现在同一次试验, 假如有朋友出现在同一次试验中, 他们将被分配到不同的团体中去。通过这种方法, 研究者使得同伴关系混淆任务或社会凝聚力的可能性降低到最小程度。

团体研究者面临的一个重要问题是所需要的团体的大小。如果团体的大小是在研究中需要加以操纵的, 那么这就不成问题。但是, 大多数研究要求团体的大小必须固定。这样, 多大的团体才是合适的呢? 团体越大, 当然需要召集的被试就越多, 这样就可能给研究的进行带来很多问题。例如, 对于一个 $2 \times 2 \times 2$ 的因素设计, 如果采用的是5人的团体, 可能就需要600~800名被试(8个单元, 每个单元上有15~20个团体), 这样, 研究者可能试图采用小的团体, 比如两人组成的团体。但是, 一些研究者指出, 两人组与三个或更多的人的团体之间存在着质的差别(Simmel, 1902, 1950)。因此, 除非依据实验设计的目的本身需要使用两人团体(例如, Zaccaro, 1984), 研究者应该使用不少于三人的团体。

另外一个重要的问题是, 在每个实验处理水平上的团体数目。在个体研究中, 通常每个处理水平上有20个以上的被试便足以稳定抽样误差(不过, 由于统计检验力的要求, 具体的数量可大可小; Cohen, 1977)。但是, 在团体研究中, 当团体作为分析单元时, 得出的平均数, 比使用相同数目的个体水平的平均数稳定。尽管团体水平的研究尚未出现一个对数据采集数目的惯例, 但是在每个处理上少于20的团体数也能够满足相对数据稳定性的要求。事实上, Nelson, Schneider (1990) 进行了一个对相同数目的个体水平和团体水平的统计效力的蒙特卡罗模拟分析。他们发现, 对于回归系数的显著性, 团体水平的分析要比个体水平的分析具有更强的统计检验效力。这样, 对同样的统计分析效力, 团体水平的研究比个体水平的研究需要的单元数目就要少一些。

自然情景

当数据是从自然情景或者已存的团体中收集时, 研究者需要考虑团体的构成和可能的团体扩展范围。与实验研究中随机分配形成的团体不同, 在自然情景中, 团体是本身存在的, 团体成员自愿选择进入, 某团体或团体所在的组织。这样的团体可能更具有一致性, 具有需要作为协变量进行统计控制的某种特质。此外, 获得一个团体研究样本, 通常需要团体或组织的领导者的同意。成功的或有效率的团体领导者通常比表现差的团体领导者更有可能同意外部研究者对其团队开展研究。这样, 研究者在自然情景中挑选所研究的团队时, 进行研究的时候需要考虑到对团队范围的限制。

已存的团体在大小和状态上不可能是完全一致的, 而这些和其他相关变量可能是大部分团体现象的重要决定因素。因此, 研究者需要确定样本中团体与团体之间在性质和特征上的差异。这些特征在随后的统计分析中可能要作为协变量加以考虑。

对团体过程的测量

在团体研究中,对主要预测变量和被预测变量的划分取决于这个研究的概念基础。但是,因为大多数自变量(输入)和因变量(输出)关系的中介变量是团体内的过程或者人际互动,多数研究者会考虑到这些过程并收集这些数据。团体过程可以通过自我报告或者观察的方法来进行测量。在一些情景中,自我报告可能是非常不可靠的,并且受到团体活动结果的影响(如团体绩效)。尤有甚者,如果团体输入和输出也是由团体内的成员来报告(如团体凝聚力、团体规范),那么任何关于输入、过程和输出变量的显著结果都可以是由于某种共同的方法偏差造成的,因而观察法通常被认为是收集团体过程数据的方法。

一些研究者使用编码者来观察和记录团体过程,问题是,除非编码者受过高度训练并且记录的方法相当简明,大量的团体互动信息很可能丧失。这是因为团体中的互动过程通常比较复杂,发生得很快,而且有多种同时发生的行为。因此,观察研究应该对团体活动进行录像和录音,以便拥有更为永久的对团体整个过程的记录,从而可以让多位评估者对这些活动进行更为仔细的分析。用以观察团体互动过程的系统有3个例子:BalesIPA(Bales, 1950a, 1950b),SYMLOG(Bales&Cohen, 1979),TEMPO(Futoran, Kelly, & McGrath, 1989)。这些观察系统尽管存在很多方面的不同,但是都提供了一套针对不同团体操作环境的编码方法。所有3个系统都区分了以任务为导向的团体互动和以社交为导向的团体互动。例如,TEMPO(Futoran等, 1989)提供了包括产出功能(任务导向)和非产出功能(社会导向)两个类别,产生功能内再分为与内容和过程有关的类别以及提出思想和评价思想的类别。

观察数据可以通过外显的或隐藏的方式进行收集。在实验室环境中,隐藏的方式包括使用单向玻璃或隐藏的摄像头。尽管单向玻璃的使用非常广泛,但是,它可能会误导团队成员,因为团队成员通常总会留意单向玻璃后是否有观察者,这样会对团体活动的测量产生干扰。隐藏的摄像机可以避免这个问题,但是可能会带来职业道德方面的麻烦。我们发现,在几个研究中,将摄像机直接安放在团体活动的设施里并没有带来显著的霍桑效应。团队成员在报告中表示,在相对短的一段时间后,他们已经习惯于摄像头的存在,因此,团体的行动不会受到它的影响。所以我们建议,在告知被试的情况下使用摄像机,这样做不但可以避免职业道德方面的问题,而且并不一定干扰团体活动。需要提及的是,为了采集到团体中所有成员的反应和互动,采用多于一台摄像机进行记录往往是必需的。

团体研究的数据分析

分析数据时要注意的问题取决于研究的基本设计。如果因变量仅仅是在团体水平上收集的（即只有团体水平的数据构成研究的数据集），那么，统计分析和统计假设与在个体水平上进行的数据统计分析是一致的。例如，在一个研究中，如果将在团体的整体水平上产生的团体绩效作为因变量，而自变量是任务导向的凝聚力和社交导向的凝聚力，对每一种凝聚力的高低进行操纵，这样一个 2×2 的方差分析就可以用来分析实验数据（Zaccaro & McCoy, 1988）。当自变量/预测变量是在个体水平收集的，而因变量/被预测变量是团体水平的数值时，团体水平的数据分析仍然是需要的。因此，在前述研究中，当预测变量是每个成员报告的对凝聚力的个人感受，而不是被研究者操纵的凝聚力水平，便可用团体成员的凝聚力感受的平均数和团体绩效进行相关分析。

但当数据是从团体成员的个体水平上收集时（例如，在层次结构设计中，个体被嵌套在团体之中），在进行统计分析之前要决定合适的统计单位，也就是说，研究者需要判断是否数据存在“团队效应”。这个效应是指个体对调查的反应受到他/她所处团体的影响。当个体在一个团体中进行互动时，他们对调查的反应不可能是相互独立的。通常（尽管并非总是如此）他们的反应会倾向于表现得更为一致。

有几种统计程序可以用来确定数据中是否存在团体效应。如果数据是在被试被随机分派到各团体中，且从其他自变量被操纵的实验中收集的，那么，个体所在的团体就可以被看作是在处理内团体嵌套方差分析随机变异的一个来源（Anderson & Ager, 1978; Myers, 1979）。如果团体效应是显著的，那么，它的均方就作为检验处理效应的误差项；如果团体效应不显著，那么用聚集的误差项来测量这些效应（参见Anderson和Ager, 1978; George, 1990; Myers, 1979; Yammarino和Markham, 1992; 对这些程序的描述）。

Kenny和La Voie（1985）建议使用组内相关来测量团体效应。这种统计方法通过比较团体内和团体间的方差来衡量在团体水平上的方差。如果团体内方差显著地比团体间方差小，则可认为团体效应是存在的。在这种情况下，Kenny和La Voie建议，同时在团体和个体水平上计算相关，从而用团体效应来对个体水平上的相关进行修正以及用个体效应对团体水平的相关进行调整（Kenny, 1985; Kenny, La Voie, 1985; 另一种计算团体效应的程序方法参见James, Demaree和Wolf, 1984）。

当数据是从团体成员的个体水平上收集时，在进行其他的统计分析之前估计团体效应是绝对必需的步骤。这种对团体效应的评估和分析即使是在那些研究假设针对个体效应的研究中也是必需的，因为在这种情况下团体效应可能是个体差异的来源。与之对应，如果一个研究者对团体现象感兴趣，却在估计团体效应时没有发现效应的存在，这

时对数据的解释就要格外小心。因为，如果个体所在的团体对于其行为毫无影响，研究者需要考虑收集到的数据是否与团体或集体行为有关。我们知道，在一些研究中，研究者旨在研究团体效应，却没有发现其存在，便转而在个体水平上对数据进行分析。但是，他们却将结果用团体的影响来解释结果！如果对数据的分析没有发现团体效应，这样的解释是没有根据的。因此，团体研究者需要注意对数据中团体效应的存在与否以及效应队随后的数据分析和解释的影响。

总 结

在本章中，我们指出了团体研究和个体研究之间的几种差异。团体研究与个体研究之间差异的存在，是因为在多数团体研究中团体中存在着个体间的交互作用。团体研究者通常会对这些交互作用感兴趣。Lewin (1951) 将团体描述为“动态的整体”，并指出：“动态整体的结构特征与各个次部分的结构特征是不同的 (p.192)。”此外，他还提到：“结构特征更多地受制于各个部分之间的关系模式而不是那些部分本身 (p.192)。”这些团体中个体之间的关系和互动应该引起研究者的格外注意。本章也描述了一些在团体研究中收集和分析数据的程序。我们相信，遵从这些步骤，研究者将会提升团体研究的水平以及获得对团体现象更为有力的理论。

推 荐 阅 读

我们为读者推荐下述读物，作为对本章中主题更为详尽和深入的探讨：Anderson和Aner的“小团体研究中的方差分析”（1978）；Driskell和Salas的“可以在设计的情景中研究真实的团队吗？团体研究在理解团队中的价值”（1992）；Forsyth的《团体动力学》（*Group Dynamics*, 1990, 第2章）；Kenny和La Voie的“分离个体和团体效应”（1985）；McGrath的《互动和绩效》（*Groups: Interaction and Performance*, 1984, 第3、4章）。

参 考 文 献

- Anderson, L. R., & Ager, J. W. (1978). Analysis of variance in small group research. *Personality and Social Psychology Bulletin*, 4, 341-345.
- Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgement. In H. Guetzkow (Ed.), *Groups, leadership, and men*. Pittsburgh: Carnegie.
- Asch, S. E. (1955). Opinions and social pressures. *Scientific American*, 193(5), 31-35.
- Bales, R. F. (1950a). *Interaction process analysis; A method for the study of small groups*. Cambridge, MA: Addison-Wesley.
- Bales, R. F. (1950b). A set of categories for the analysis of small group interaction. *American Sociological Review*, 15, 257-263.

- Bales, R. F., & Cohen, J. S. P. (1979). *SYMLOG: A system for the multilevel observation of groups*. New York: Free Press.
- Cartwright, D. (1968). The nature of group cohesiveness. In D. Cartwright & A. Zander (Eds.), *Group dynamics: Research and theory* (pp. 91-109). New York: Harper & Row.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Cohen, M. (1993). The naturalistic basis of decision biases. In G. Klein, J. Orasanu, R. Calderwood, & C. E. Zsombok (Eds.), *Decision making in action: Models and methods* (pp. 51-99). Norwood, NJ: Ablex.
- Driskell, J. E., & Salas, E. (1992). Can you study real teams in contrived settings? The value of small group research to understanding teams. In R. W. Swezey & E. Salas (Eds.), *Teams: Their training and performance* (pp. 101-124). Norwood, NJ: Ablex.
- Duffy, L. (1993). Team decision-making biases: An information processing perspective. In G. Klein, J. Orasanu, R. Calderwood, & C. E. Zsombok (Eds.), *Decision making in action: Models and methods* (pp. 346-359). Norwood, NJ: Ablex.
- Forsyth, D. R. (1990). *Group dynamics* (2nd ed.). Pacific Grove, CA: Brooks/Cole.
- Futoran, G. C., Kelly, J. R., & McGrath, J. E. (1989). TEMPO: A time-based system for analysis of group interaction process. *Basic and Applied Social Psychology*, 10, 211-232.
- George, J. M. (1990). Personality, affect, and behavior in groups. *Journal of Applied Psychology*, 75, 107-116.
- Gualtieri, J., Parker, C., & Zaccaro, S. J. (1995). *Group decision making: An examination of decision processes and performance*. Manuscript submitted for publication.
- Hanges, P. J., Nelson, G. L., & Schneider, B. (1990). *Levels of analysis and statistical power*. Paper presented at the annual meeting of the American Psychological Association, Boston.
- James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology*, 69, 85-98.
- Janis, I. L. (1982). *Victims of groupthink* (2nd ed.). Boston: Houghton Mifflin.
- Kenny, D. A. (1985). The generalized group effect model. In J. Nesselroade & A. von Eye (Eds.), *Individual development and social change* (pp. 343-351). New York: Academic Press.
- Kenny, D., & La Voie, L. (1985). Separating individual and group effects. *Journal of Personality and Social Psychology*, 4, 339-348.
- Latane, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology*, 37, 823-832.
- Lewin, K. (1951). *Field theory in social science*. New York: Harper & Bros.
- Lipshitz, R. (1993). Converging themes in the study of decision making in realistic settings. In G. Klein, J. Orasanu, R. Calderwood, & C. E. Zsombok (Eds.), *Decision making in action: Models and methods* (pp. 103-137). Norwood, NJ: Ablex.
- McGrath, J. (1984). *Groups: Interaction and performance*. Englewood Cliffs, NJ: Prentice Hall.
- Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, 69, 137-143.
- Milgram, S. (1965). Some conditions of obedience and disobedience to authority. *Human Relations*, 18, 57-75.
- Montgomery, H., & Svenson, O. (1989). *Process and structure in human decision making*. New York: John Wiley.
- Myers, D. G., & Lamm, H. (1976). The polarizing effect of group discussion. *American Scientist*, 63, 297-303.
- Myers, J. L. (1979). *Fundamentals in experimental design* (3rd ed.). Boston: Allyn & Bacon.
- Orasanu, J., & Salas, E. (1993). Team decision making in complex environments. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsombok (Eds.), *Decision making in action: Models and methods* (pp. 327-345). Norwood, NJ: Ablex.
- Pennington, N., & Hastie, R. (1986). Evidence evaluation in complex decision making. *Journal of Personality and Social Psychology*, 51, 242-258.
- Roethlisberger, F. J., & Dickson, W. J. (1939). *Management and the worker*. Cambridge, MA: Harvard University Press.
- Simmel, G. (1950). The significance of numbers for social life [excerpts]. In K. H. Wolff (Ed.), *The sociology of Georg Simmel*. Glencoe, IL: Free Press.
- Tversky, A., & Kahneman, D. (1972). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 4, 207-232.
- Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, 211, 453-458.
- Wallach, M. A., Kogan, N., & Bem, D. J. (1962). Group influence on individual risk taking. *Journal of Abnormal and Social Psychology*, 65, 75-86.
- Yammarino, F. J., & Markham, S. E. (1992). On the application of within and between analysis: Are absence and affect really group-based phenomena? *Journal of Applied Psychology*, 77, 168-176.
- Zaccaro, S. J. (1984). Social loafing: The role of task attractiveness. *Personality and Social Psychology Bulletin*, 10, 99-106.
- Zaccaro, S. J., & McCoy, M. C. (1988). The effects of task and interpersonal cohesiveness on performance of a disjunctive group task. *Journal of Applied Social Psychology*, 18, 837-851.
- Zajonc, R. B. (1965). Social facilitation. *Science*, 149, 269-274.

第四部分

数 据 分 析



第 13 章

数据预处理及初步分析

Stephen J. Dollinger

David L. Dilalla

数据预处理

实际上,不仅在心理学研究中,而且在所有科学研究和寻求真理的过程中,都需要尽量避免错误的发生。尽量避免错误是心理学研究者必做之事,比如从重新检查数据到显著性检验以避免Ⅰ型错误,从采用有足够统计效力的实验设计以避免Ⅱ型错误到思考如何来解释自己的(或者他人的)数据。既然大部分心理学数据是由研究助手来收集的,其是否达到或接近无误的数据就成为心理学家们评价一个助手能力高低的标准。不论是对单个科学研究、临床个案研究,还是整个领域的整合,都要求我们能尽可能地让数据接近有效并且无误。初步分析过程能够通过识别几种错误来源而达到这个要求,并且能够帮助研究者更好地理解数据的深层含义。

数据预处理

Smith、Budzeika、Edwards、Johnson和Bears(1986)发表了一篇叫做《预处理指南》的文章,这篇文章提供了如何预处理数据的有用信息。Smith等人提出了一系列有效预处理数据的准则,尤其是针对在二级分析中用到的大量数据。Smith等人的建议适用于很多数据处理的情况。总结一下他们建议中的共同主题就是:充分注意细节!

数据搜集之前和搜集的过程中

Smith等人(1986)提出,如果我们不能明确指出对标准量表进行了怎样的改变,标准的量表就不应该被轻易地变更。同样,研究者在设计如何测量被试的反应时,应该对接下来的分析方法成竹在胸。如果不这样做的话,一般情况下就会增加本可以避免的复杂性。Smith等人指出,有些研究者为了赶在最后期限之前完成任务,往往在问卷中引进了一些后来不知道如何对待的项目,这种情况很可能是因为他们对量表做了不合适

的改变。

一般说来,在把一篇学位论文提交给杂志发表之前,尽管对数据再次分析(用不同的方法)是很普遍的,但在搜集之前就计划好数据的分析,始终都是十分明智的(Yaffee,本书的第15章)。这样做可以对可能产生的实验研究结果有更加深入的分析。

(最近的一个实验研究在这方面做得颇为欠缺,一个学生意识到他的理论假设由于没有很好地考虑如何测量而导致无法验证,并且这个问题是直到所有的数据都搜集完了,要开始进行统计分析时才被发现的。)这里强调的是一旦数据搜集完了,人们就有了所谓的事后智慧,知道这个实验研究本应该如何设计。对实验研究如果有预先的数据分析计划,将有助于研究者获得这样的智慧。如果时间和其他条件都允许的话,最好能做一个预研究。

Smith等人(1986)还提出,研究者要积极地参与数据搜集工作——而不要把这个工作完全交给研究助手——而且研究者要检查各种实验工具,以防止疏漏。

数据搜集之后

Smith等人(1986)建议,研究者应该相当了解他们使用的测量工具以及计分标准和整个过程。这些应该在整个计分过程结束之前就被考虑到。当数据搜集完成之后,应当保证所得到的各个项目或者量表的数据与计分标准吻合。如果反应是以1—2—3—4—5的格式编码,但其中的某些项目是反向编码,这些应当体现出来(比如,重新编码这些项目,或者在这里用6来减去这些数值)。如果不能正确地重新编码反转的项目,那么本来可能很好的数据将变得一文不值。考虑到今后的抽样调查,最好不要改变原始的变量,而是重新定义新的变量,来代表被重新编码的相应变量。另外,知道了可能的域值之后,对那些超出可能域值范围的数据要谨慎。频率分布表和分布图可以有效地帮助你识别那些异常值。

Smith等人(1986)还提出了另一个建议,这对刚刚步入研究领域的新手来说尤其具有参考价值,那就是:“把那些测量相同特征的项目按照同样的计分方向来重新编码。”比如,测试的项目是用来测量焦虑和抑郁的,如果两个量表中每个项目上高的分数代表更高程度的负性水平,那么解释起来的话就会更加容易,并且也不会出错。进一步来说,研究者干脆就把这样的编码方式——即用高分值代表所研究的特性的更高程度或更高水平——当作一种习惯。如此一来,如果研究者测量的是外向——内向特质,研究者与被试的工作就变得相当简单了(如果使用诸如A、B、C这样的字母,那么,在使用的过程中,A应该表示更加让人满意或更好的方面,因为学校中A代表的是更好的成绩)。

数值缺失在不同的统计软件包里是用不同的方式来编码的。在有些软件中,专门有一个特定的数字(如9)是用来表示缺失的数值。也有些软件是用一个小数点或者空白来

表示数值缺失。在进行数据分析之前要了解这些情况，因为如果忽视了这一点，很有可能两个含有缺失数值（如9）的变量，会由于人的不谨慎而造成错误的显著相关（作者的一个研究助手就曾经经历过这样的尴尬）。为了避免产生这样把缺失数值错当成有效数值的情况，还有一种有效的方法就是尽量去选择那些不太可能成为有效值的数字（如999）。另外，很关键的一点就是要在统计分析软件中定义好用什么值来代表数值缺失，以避免那些缺失数值和有效数值混淆，从而影响统计分析结果的正确性。最重要的就是要了解你所使用的统计软件（如果你使用小数点来代表数值缺失，那么，一定只使用单个的小数点，即使这个变量分配了两个字符以上的位长）。

Smith等人针对几个数据集合的汇总提出了建议。数据的汇总无论是在初级分析还是在次级分析的情况下都有可能被用到。数据的输入顺序很重要；如果一个被试有不只一条的记录，而其中有些被试的记录行被删除了，那么，这样的统计分析结果将根本无法解释。或者更糟糕的是，统计分析结果似乎可以解释，但却是错误的解释。当输入的项目和数据记录不吻合时，同样的问题也会发生。重新在那些描述性统计表格中寻找异常值，或者打印出某些被试数据的列表，与原始数据进行对照。当几个数据集合汇总成一个大的样本时，一定要保证它们具有可比性，并且是按照同一个变量排序的（如被试编号）。另外要注意同一个被试编号只能用于一个被试！

还有一种做法，也可以保存原有的若干个原始数据文件（每一个文件里面都有一些相同的识别号码），新建一个“工作”分析文件（在spss的术语里面叫做系统文件），然后把这些原始数据文件汇总到这个主分析文件里面。这样做避免了逐个汇总所有数据记录的需要。举个例子来说，在一个我们经常用到的较大的数据集合里，对诊断数据、MMPI数据、使用酒精数据、人格特质数据等，都有不同的系统文件，每个文件都有一个共同的ID变量。在新建一个工作的数据集合时，只要从中选择我们所感兴趣的变量（通过ID来匹配），来建立这个新的系统文件就可以了。

Smith等人（1986）在总结部分提出了对变量进行强迫性标签的建议。这包括要备份测量工具、计分的标准、对变量的编码（例如：控制组的被试是用1来表示，实验组用2来表示）。对变量要清楚地标签，以保证即使是把数据交给一个陌生人，他也能凭借变量的标签看懂。如果你把这个项目搁置一个月，以完成别的任务，这个数据的很多方面都会被遗忘。除了要把计算机里的文件标注之外，还要保存一个分析日志，日志里面记录对数据进行了怎样的处理，这些都能帮助研究者回忆，而且即使是不只一个人在进行数据分析的时候，也能做到有条不紊。

发现“糟糕”的被试

在一个关于人类被试的讨论经典中，Schultz（1969）指出，在心理学研究中，很多被试都试图成为一个好的被试，或者说一个可信的被试，为研究提供有意义的数

助实验者达到实验的目的。这里，我们所担心的是不可信的或者说那些糟糕的被试——他们或者是无意或者是有意地不认真做反应（例如，带着“早点儿结束这个烦人的实验”这种想法，不认真去看实验的要求和实验的各个项目）。这样的被试往往在不同项目上有一连串相同的反应（例如：太多连续的“5”）或者有相同的反应模式（例如：1-2-3-4-5-4-3-2-1），或者对测试的项目以半随机的方式来反应。

在数据收集的过程中，记录被试完成实验的时间是十分有必要的。可以考虑删除那些完成实验太快的被试的数据。如果你是在测试一个较大的样本中进行团体测试（Zaccaro & Marks, 本书第12章），还有一个方法就是按照被试完成时间来标记每一个被试，看看这个变量是否跟我们感兴趣的测量有关。正如参加的一个学期的学习所花的时间能够反映人格和被试的投入一样（Roman, Moskowitz, Stein, & Eisenberg, 1995），完成一个实验所花的时间也能作为一个影响研究的内部效度的指标。

很显然，在问卷中，特别是那些以光学扫描形式审阅的问卷中，靠视觉检查就可以发现一些可疑的糟糕被试。如果研究助手把问卷中的数据打孔了的话，一些异常的反应模式就比较容易注意到。

数据中的异常值应该易于被检测到。比如对性别这个变量进行编码（男=1，女=2），一个频率分布图就很容易发现那些因为种种原因而不小心被插入的其他数字。如果这些异常值在对性别和其他变量做相关分析之前没有被觉察到，例如有些被试的性别是3、4、5这样无意义的数字，可以想像一下做出的相关该是多么的具有误导性。

如果所有的数据都存储在计算机里面，还有一个有用的方法也可以考虑。新建一个变量，表示一个被试在所有项目上的反应情况（不考虑其中的量）。具体来讲，第一步是新建一个变量，比如Grandmean和Grandsd。我们使用这些名称来表示被试在所有项目上的总体平均值和总体方差。接下来，找出那些超出Grandmean和Grandsd正负3、4或更多标准差之外的被试。分析了几个不同的数据集合之后，我们发现，用这样的被试列表可以确定那些连续使用高的分数来评价项目的被试，这样的反应会导致在Grandmean上的高分值以及在Grandsd上的低分值。那些使用1-2-3-4-5-4-3-2-1反应模式的被试会在Grandsd上表现出可疑的高分值。

还有一个有用的办法，就是在数据集合中建立几个验证项目。比如，插入一些诸如“我的衣服和鞋子是自己做的”（Jackson, 1984）这样通常不可能发生的事情。如果被试的回答是肯定的，那就可以确定这个被试的回答是随意性的。不过，过多使用这样的验证项目会引起一些效应，被试会认为他们的参与不是很严肃而是一件很随便的事情。不同版本的大五NEO人格问卷（Costa & McCrae, 1992）都包括这个简单的项目“我尽量让自己诚实准确地回答问卷中所有的问题”。如果有被试在回答这一项的时候是选中性或者不同意，那要么是他或她根本就没有去读这一项，要么是把5点量表的正性和负性两端弄反了。根据这种策略，高明的研究者在一个长的问卷中也包括进这种项目，但目

的是为了识别相反方向的反应；比如：“我费了很大的力气去弄懂这个量表中的许多单词。”对这一项认可的回答，尤其是那些非常同意的回答，往往表明被试或者没有好好读这一项，或者实际上没有弄懂问卷里大多数的问题。综合这种事前判断“糟糕”数据的方法以及事后Grandmean和Grandsd方法，我们就能识别哪些数据在主要统计分析进行前应该被删除。

以上所说的方法对于单变量的奇异值是极其有效的，但对于多变量的奇异值来说，就需要更多的方法。几个联合起来识别问题数据的处理步骤可能更加有效（Comrey, 1985; Barnett & Lewis, 1978; Hawkins, 1980）。Comrey的方法（1985）是基于特定被试相对于均值的平均方差。Comrey指出，不能因为一个公式算出来被试不合格就轻易地把被试的数据删除，相反，所有这些识别奇异值的方法所识别的仅是那些需要仔细检查的个案。在使用Grandmean/Grandsd的基础上，我们认可这种做法，很有用的一个步骤就是设计一系列数据的初步图表，这些图表也能帮助识别奇异值（Anscombe, 1973）。

“糟糕的”主试

从糟糕的被试，我们转过来讨论糟糕的主试，这些人也可能糟蹋数据。在促进理解心理实验以及实验者效应这些方面，Rosenthal的大量工作（1966, 1944）做出了重要的贡献。Barber（1976）通过综合、评论已有文献，总结出人类研究的十大缺陷。其中的一项就涉及我们自身。不过首先我们简单地评论一下主试未遵循实验程序所产生的效应和主试捏造数据的后果。我们假定本章的读者都是有良心的诚实人，知道遵循实验程序、不捏造任何数据，是作为一个研究助手所应具备的道德要求。篡改和捏造数据是极为错误的，它违背了整个科学的理念，会中止一个学者的职业生涯（注意，删除数据的奇异值不是一个道德的问题；相反，在一些研究领域中，比如涉及反应时的实验，根据大家所认可的标准来如此处理数据，这是一种常识）。

说了这些之后，我们来谈论在道德问题上轻微一些但更加实际一些的主试记录错误所产生的后果。很多研究（Barber, 1976）表明，主试在没有恶意的情况下，也确实会错误地记录数据。人们确实会犯一些“无心的”错误。在Laszlo（in Rosenthal, 1966, p.13）的一项研究中，3位主试测试了64名被试，结果发现，这些主试在记录时犯错误的比率分别为6%、22%和26%！另一个研究是Rosenthal等人（1964）做的，其中30名主试参加了数据抄写工作。在3 000个数据点中，有20个抄写错误，总错误率仅为0.67%。从这些研究中，我们能够得出什么信息？人是会犯错误的，不过错误也是可以通过重复检查发现的。从应用行为分析的方法学中我们可能获得一些启示，数据的收集者们（比如行为观察者）如果知道他们的可信度将在整个实验研究中被定期检查，那么，他们的工作将会做得更加认真和精确。

如果主试的任务不仅仅是做一些记录性的工作，我们还得警惕会出现实验者预期效

应的可能性——最经典的就是教室里的皮格马利翁效应了。临床领域的研究生可能会碰到这一问题。考虑一下在进行心理测试的时候，一个人的预期会对结果有多大的偏差效应，在临床领域这种效应是比较大的（Badad, Mann, & Mar-Hayim, 1975; Sattler, Hillix, & Neher, 1970）。

初 步 分 析

收集完数据并且去除了那些错误和奇异值之后，做初步的分析仍然是十分有用的。实际上，我们通常会建议说，论文的“结果”部分应该遵循这样的格式：对“初步结果”的简单描述、“主要的结果”以及“补充的结果”。初步的分析能帮助研究者确定如何更好地继续下面对数据内涵的描述。初步分析首先应是考虑这个样本是否“有效”——也就是说，这些描述性的分析（ M , SD ）和测量的可信度（如 α 系数）是否与其他研究者所报告的结果具有可比性。显然，这里所关注的是实施测试时的标准或者说控制的条件，而不是是否具有处理效应。如果你的一些基本描述统计量失效了（比如，一个大学生样本的平均智商是85），那你的被试就不具有代表性，他们可能误解了你的指导语，或你可能在简化数据时犯了错误。

在初步分析阶段之外，你可能希望忽略其中一些变量。在这里，你的目的可能是表明主要的结果是不受诸如年龄、性别、年级等因素限制的。这样，你就希望检验性别造成的差异、其他变量和性别的交互作用、由性别所引起的不同相关模式等。如果使用的是真实验设计而不是相关设计，你还可能建立一些检验实验操作效果的问卷项目，用来报告处理的内部效度。除了初步分析的这些情况之外，当你使用繁多的项目或量表时，还可以通过因素分析来简化数据。

使用因素分析

有许多特定的初步分析跟因素分析有关。在把数据提交给高速计算机和简单易用的统计软件包处理之前，研究者在充分进行预见和计划的情况下会把数据做因素分析。目前很常见的就是研究者很有兴趣把数据糅进去进行因素分析，作为对数据进行广泛性探索分析的一部分。鉴于因素分析能够从任何可靠的数据中抽取因素，我们有必要认识到，某些数据的局限性可能导致错误的解释。

正确使用因素分析的第一步当然就是相关矩阵的视觉检查。如果各个变量之间的相关普遍都非常小，这个矩阵里就不可能存在一些可观察到的共同因素。经过最初的检查之后，有必要去评估一下很多因素分析软件都会提供的两个基本诊断统计量。这些信息有助于判断这些数据是否适合因素分析。

Bartlett的球形检验提供了视觉检查的一个统计评估指标，它检验了将要进行因素分

析的变量的相关矩阵是否真的是单位矩阵（矩阵的对角线上所有的元素都是1，对角线之外的其他元素全都是0）。用来检验的统计量是符合卡方分布的。当检验统计量的值足够高的时候， p 值达到统计学上的显著水平，研究者很容易地拒绝虚无假设，即相关矩阵是单位矩阵。一般来说，球形检验是个相当宽松的检验，球形统计量的显著值对于进一步进行因素分析来说是必要的，但不是充分的。

除了Bartlett的检验之外，还应该参考关于取样充分性的Kaiser-Meyer-Olkin检验（KMO; Kaiser, 1974）。这个统计指标反映的是，在多大程度上共同因素解释了变量之间的相关，是用各个变量之间两两相关系数的平方和除以相关的平方和加上偏相关的平方和。

偏相关越接近0，共同因素就越能够解释变量中增加的变异；当共同因素模型适合数据时，KMO统计量将比较高。KMO统计量偏小，表明两两变量之间的相关不能被共同因素所解释。KMO统计量的取值范围在0~1之间。Kaiser（1974）提出，KMO统计量大于0.9时效果最佳，大于0.8时效果不错，大于0.7时效果尚可，0.6时效果比较糟，0.5或以下时就不适合做因素分析。当统计量的值低于0.5时，研究者应重新考虑对观察到的相关是否具有因素模型的适宜性。还应提到的是，即使在KMO统计量表明一个因素模型不合适的环境下，因素分析的程序还是会很教条地提取和旋转它认为最适合的因素解决方案。

决定，决定

当确定数据非常适合做因素分析之后，研究者面临着无数的选择：如何进行因素的提取？提取多少因素？是否进行旋转？如果需要旋转，该使用何种类型的旋转策略？所有的这些决定都应该服从于研究的目的。

因素提取

详细讨论因素提取的方法并不属于本章讨论的范围（详见Harman, 1967, Kim & Mueller, 1978），但几个普遍的注意点还是值得提及。当分析的目的是为了把大量的变量简化为一定的易于分析的数目，主成分分析或许可以成为一个提取因素的方案。严格地说，主成分分析不是因素分析，但与因素分析的过程极其相似，在类似SPSS的统计软件包中，主成分分析常常作为默认的选项。主成分分析是通过线性变换，将原来的多个指标组合成相互独立的少数几个能充分反映总体信息的指标。第一个主成了解释了最多的变量变异，第二个次之，以此类推。最终的结果是把相关的变量转化成相互独立的主成分。第一主成分常常被研究者们用来作为结构指标，代表相关变量之间共同的变异成分。比如，一个人做了5个特定认知能力的测试，第一个从5项能力之间的相关关系中提取出来的主成分被认为是对普遍能力的测量。在“真正的”因素分析的算法中，主轴分析也用得很普遍。主轴分析所运用的线性变换与PCA很类似，不过使用的是迭代的程

序，对公因子方差比（原变量的变异中由公因子决定的比例）是被嵌入到分析相关矩阵的对角线中去的。

应该提取多少因素才足够？

因素分析中一个具有挑战性的问题是确定最合适的、应提取和旋转的因素数量。如果分析者选择软件默认设置，统计软件有时能让粗心的分析者犯错误。最常使用的默认标准是选择特征根（多少变异能被因素解释的一项指标）大于1.0的因素。另外，研究者还应参考累积贡献率，碎石图可以提供特征根大小排列的情况。一般认为（Cattell, 1966），当曲线有明显的中断时，就不应该再提取下面的因素了。最后，对于探索性的因素分析，研究者应提取比统计软件默认的选项多或者少一些的因素，这让研究者可以衡量哪种提取会更好。提供“简洁结构”（少量的几个变量，但载荷却相对很高）。

是否进行旋转

如果研究者的目的只是为了使上面所提到的数据简化，那不经过旋转的第一个主成分就可以令人满意地达到要求。然而，研究者常常有兴趣去确定对一组变量最有意义的、解释最好的结构，以便更好地了解一个特定测量的结构效度。在这种情况下，各种因素旋转的方法能够更好地解释最初未旋转过的因素结构。简单说来，旋转反映的是在参考轴上，因素模型中信息量在不同因素上的分布。在某个轴上聚集在一起的变量被认为是彼此相关的。最主要的两种旋转是正交旋转和斜交旋转。正交旋转的原则是因素之间互相独立（参考轴是相互垂直的），而斜交旋转允许因素之间存在相关（参考轴的交角可以不是直角）。在某些情况下，选择相关还是非相关的因素解决方案必须有相应的理论上的重要依据。如果没有这样的基本理由，那最谨慎的做法就是使用斜交旋转，因为行为研究者们遇到的大部分结构并不都是相互独立的。在探索性因素分析中，应综合使用正交旋转和斜交旋转，探索不同的途径，对从观察数据中提取的因素进行解释。

总结：整理数据需要吃一定的苦

作者在研究生阶段曾经有一位令人难忘的老师，他是一个教授统计学的老师。他喜欢这样提醒学生们：“不要害怕在数据面前把你的鼻子弄脏。”看上去似乎是开玩笑，但实际上这样确实能最好地保证数据的整洁性。总的说来，初步分析是最容易对数据上做统计手脚的过程。你自然想尽可能快地验证自己的假设是否被支持，但重要的一点是，记住对新数据的第一次分析（或是对老数据的重新分析）应该被看做是初步的。随着对细节问题，诸如不同变量下或项目上被试的数目，越来越熟悉，你将能敏锐地发觉数据中存在的问题或矛盾之处，也就更加能游刃有余地把握整个研究。

推 荐 阅 读

读者如果想对整理数据和初步分析有更多的了解,可参考Anscombe (1973)、Kim & Mueller (1978) 和Smith et al. (1986) 的著作。

参 考 文 献

- Anscombe, F. J. (1973). Graphs in statistical analysis. *American Statistician*, 27, 17-21.
- Badad, E. Y., Mann, M., & Mar-Hayim, M. (1975). Bias in scoring the WISC subtests. *Journal of Consulting and Clinical Psychology*, 43, 268.
- Barber, T. X. (1976). *Pitfalls in human research*. New York: Pergamon.
- Barnett, V., & Lewis, T. (1978). *Outliers in statistical data*. New York: John Wiley.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavior Research*, 1, 245-276.
- Comrey, A. L. (1985). A method for removing outliers to improve factor analytic results. *Multivariate Behavioral Research*, 20, 273-281.
- Costa, P. T., & McCrae, R. R. (1992). *The Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.
- Harman, H. H. (1967). *Modern factor analysis* (2nd ed.). Chicago: University of Chicago Press.
- Hawkins, D. M. (1980). *Identification of outliers*. New York: Chapman & Hall.
- Jackson, D. N. (1984). *Personality Research Form manual* (3rd ed.). Port Huron, MI: Research Psychologists Press.
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39, 31-36.
- Kim, J. O., & Mueller, C. W. (1978). *Introduction to factor analysis*. Beverly Hills, CA: Sage.
- Roman, R. J., Moskowitz, G. D., Stein, M. I., & Eisenberg, R. F. (1995). Individual differences in experiment participation: Structure, autonomy, and time of the semester. *Journal of Personality*, 63, 113-138.
- Rosenthal, R. (1966). *Experimenter effects in behavioral research*. New York: Appleton-Century-Crofts.
- Rosenthal, R. (1994). Interpersonal expectancy effects: A 30-year perspective. *Current Directions in Psychological Science*, 3, 176-179.
- Rosenthal, R., Friedman, C. J., Johnson, C. A., Fode, K., Schill, T., White, R. C., & Vikan, L. L. (1964). Variables affecting experimenter bias in a group situation. *Genetic Psychology Monographs*, 70, 271-296.
- Sattler, J. M., Hillix, W. A., & Neher, L. A. (1970). Halo effect in examiner scoring of intelligence test responses. *Journal of Consulting and Clinical Psychology*, 34, 172-176.
- Schultz, D. P. (1969). The human subject in psychological research. *Psychological Bulletin*, 72, 214-228.
- Smith, P. C., Budzeika, K. A., Edwards, N. A., Johnson, S. M., & Bearse, L. N. (1986). Guidelines for clean data: Detection of common mistakes. *Journal of Applied Psychology*, 71, 457-460.



第 14 章

定 性 分 析

Pamela S. Highlen

Heather C. Finley

Denzin和Lincoln（1994）把定性研究概括地定义为一种解释性的多元方法，用来研究处于自然环境背景中的人。经验的数据来自于个案研究、访谈、观察以及历史的、相关的和视觉的资料，这些数据被拿来来进行系统的研究。研究者们努力尝试理解并解释人们生活中对日常或困难事务所赋予的意义。除了实证主义的范式之外，定性研究强调：（a）无法从量和强度的角度来严格测量的一些过程和含义；（b）由社会构造的现实的本质；（c）研究者和研究对象之间的关系；（d）影响调查研究的情境约束。

定性研究范式

Guba和Lincoln（1994）指出：“方法问题和数据分析的适应性次于研究范式的问题，或者说次于研究者所持有的基本信念和世界观，这些不仅仅体现在研究者对方法的选择上，还体现在本体论和认识论方面。”因此，思考定性研究中最基本的范式问题是任何讨论定性数据分析问题的最重要的先决条件。研究者提出了很多范式分类方法（比如，Denzin & Lincoln, 1994; Guba & Lincoln, 1994; Habermas, 1971; Lather, 1994a）。由于篇幅所限，这里简单介绍5种主要的范式：（a）实证主义的；（b）后实证主义的；（c）解释性/建构性的；（d）评论性的；（e）后结构性的。下面将概述这些范式，描述的重点将放在每种范式下评价研究性质的标准上。这些标准常常被称为“认识论效度”，或者说“一系列涉及知识、由知识派生的产品和表征的规则”（Lincoln & Denzin, 1994, p.578）。“认识论效度”不仅对一个定性研究的设计和实施很重要，对接下来分析收集的数据也同样如此。

鉴于一些原因，本章重点不在于第一个范式——实证主义的，而在于后面4个范式：后实证主义的、解释性/建构性的、评论性的、后结构性的。首先，很多定性研究者相信，现实是具由社会建构的，这与实证主义的立场背道而驰。其次，很多定性研究者都

是使用后面提到的4个范式中的某一个来进行研究。实证主义范式也会简单地提一下，主要是为了说明定性研究可以在这样的范式中进行。

实证主义范式

实证主义范式的目的是获得具有预测性和支配性价值的解释。从实证主义的角度来说，定性研究是假定有一个客观的现实存在，这个客观现实既可以被验证也可以被量化。实证主义运用内部效度、外部效度、信度和客观性来评价定性研究。例如，如果研究的问卷是可自由回答的问题，对被试的回答将进行内容分析，由几个评分者独立地把被试的答案划分类别。这样就需要计算评分者之间的一致性和信度，以确保所划分的类别具有一致性。分类过的数据以频数和百分比的形式给出，以提供对研究假设的确证性支持或为将来进一步的定性研究指明方向。另一个在实证主义范式下进行定性研究的例子就是对心理咨询过程的研究。根据一些被确认维度（比如，咨询师的类型以及咨询对象的言语陈述），将治疗过程撰写出来并且进行内容分析，使用描述性统计或其他一些统计方法，对数据加以分析。在实证主义范式下的定性研究遭到了批评，比如它的内容分析缺乏严格性（Guba, 1981），它抛弃了所收集数据中隐含的意义，把一些普遍性的数据不恰当地应用到特异性的个案中。

后实证主义范式

后实证主义范式和实证主义范式具有相同的目的：获得具有预测性和支配性价值的解释。不过，后实证主义研究者努力地想解决实证主义范式遭遇的批评。客观性的现实仍然被认为是存在的，但是这个现实只能被大约地估计〔把这种观点称为“批判现实主义”（Cook & Campbell, 1979）〕。尽管强调对理论和发现的可验证性，但它也还是会使用多种不同的方法。研究者试图寻找一个客观的立场，但是他们明白研究者和研究对象之间的交互作用还是会影响数据的。情境理论（Grounded-theory, Strauss & Corbin, 1990）是使用得最为广泛的后实证主义范式设计，因为它提供了一套明确的步骤，与实证主义科学紧密相连。认同情境理论的研究者从数据中构建理论；如果相关的理论已经存在，则根据新的数据来修正这些理论。比起其他定性研究范式的方法来，受实证主义传统影响的研究者更加容易接受和采用这种演绎方法的建立理论。面谈访问和观察常常是情境理论研究的数据来源。研究者从多个侧面来进行概念性的解释。“编码过程——包括很重要的一致性比较过程，涉及理论的提问、取样、概念的延展以及它们之间的相互关系——能够避免研究者接受他们自己立场上的观点，在某种程度上迫使研究者对自身的观点接受质疑和修正（Strauss & Corbin, 1994, p.280）。” Strauss和Corbin（1994）修订了针对定性研究的“好科学的规则”，包括了意义、理论—观察的兼容性、可泛化性、一致性、重复性、精确性以及可验证性。尽管有些模糊，这些原则仍然被用来评判

后实证主义定性研究的质量。Rose和Jevne（1993）对运动创伤的心理过程研究是一个相当好的情境理论研究的事例。

解释性/建构性范式

解释性/建构性范式的主要目标在于理解被试的个人世界。现实是通过人际交往建构的，并被看成是相对的（不存在一个单一的外部客观事实），所以就存在多个现实。随着研究的展开，研究者和研究对象之间的互动创造出调查结果。这个过程是反射性的；来自研究对象的信息一样将反馈给他们，以证实信息的精确性（成员核对）。常常使用“详尽描述”（Geertz, 1973）这种手法，通过呈现“把人与人互相联系起来的细节、背景、情感和社会关系网”，把被试的生活背景和意义揭示出来（Denzin, 1989, p.83）。详尽描述，允许读者对结果形成自己的解释（Patton, 1990）。在解释性/建构性范式下进行的调查研究所使用的方法能够让研究者精确地记录对被试的观察结果，并揭示被试赋予他们生活经验的意义（Denzin & Lincoln, 1994）。

建构主义，正如Lincoln和Guba（1985）所提出的那样，提供了在解释/建构范式下进行定性研究的系统方法。研究者作为一个研究的推动者和参与者，目的是理解并进一步重新建构个体的各个方面（对意义的解释）。她/他一方面寻求意义的一致性；一方面在信息和意识改变时，产生新的解释。最初，研究者提出“可信赖性”可以作为一套评判研究质量的标准（Lincoln & Guba, 1985）。可信赖性包括4个部分，与实证科学的方法论标准相对应：（a）可信性（内部效度）；（b）可迁移性（外部效度）；（c）可靠性（信度）；（d）确信性（客观性）。通过在一个领域中长期的研究、坚持不懈地观察、三角测量（对来源、方法和研究者）、同行报告、对反例的分析、持续的主观考察（实时监测研究者对数据的假设和解释）以及成员检查，可信性可以得到提高。使用详尽的描述能够为可转移性奠定基础。而使用相互交叉的方法和可靠性审计（对调查过程中的变化记录）能提高可靠性。确信性依赖于跟踪监测（要保留研究中的多种记录，比如书面的现场记录、备忘录、现场日记、过程记录、个人随笔以及对反馈的记录）。

最近，Guba和Lincoln（1989）又提出了一个高于可信赖性标准之上的真实性标准。他们主张，在评价定性研究时，结果、成果以及商谈的过程与方法论的标准（如可信赖性）同等重要。可确证性标准包含5部分：恰当性（对各种结构和基础价值观的依赖程度以及侧重程度）、本体论上的真实性（个人的构念多大程度上被扩展并精细化）、教学上的真实性（被试对他人构念的理解在多大程度上提高了）、催化效果真实性（多大程度上被试被激励去行动）以及策略上的真实性（参与者在多大的范围内有行动权）。真实性解决了评论性和后结构主义范式提倡者所关注的伦理道德性和授权性问题。

评论性范式

评论性范式的目的是解放和转型——让被试能够获得必要的知识，有能力控制和支配自己的生活（Gibson, 1986）。评论性的研究是多方的、合作性的，并依赖于被试（常常被称为合作调查者）的经验。它是由解释性理论组织起来的，比如女权主义、新马克思主义、参与性调查。尽管很多定性研究者都知道，不存在没有价值判断的调查（Janesick, 1994），评论性范式的研究者公然地把自己的价值判断和理论立场带入研究过程中。他们和合作调查者进行问答式的对话，以便把意义赋予更广阔的社会效力、统治和历史等结构（Lather, 1994a）。正如Guba和Lincoln（1994）所提出的那样，这种“对话必须在本质上是辩证的，能够把无知和误解（认为历史中介的结构是一成不变的）转化为更加见多识广（能够看到结构如何产生变化，理解造成这种变化的行为）”（p. 110）。

对评论性研究的评价是基于多大程度上的：（a）社会、政治、伦理、经济、文化以及性别这些元素在研究的过程中有所考虑；（b）调查消除了无知和误解；（c）调查产生了实际的改变现有结构的行为（Guba & Lincoln, 1994, p. 114）。Jones（1989, Lather曾引用过, 1992）研究新西兰在校的青少年生活时，考察了种族、阶级和性别的交互作用，运用的就是评论性范式。她使用了新马克思主义的社会理论，通过批评人种学创立了自身解放的理论，促使个体能够与压迫的环境斗争。

后结构主义范式

后结构主义范式的主要目的在于解构。也就是说，动摇和挑战任何把社会结构现实看成是完整知识的观点。解构是建立在两个原则基础之上的：“情境决定意义以及情境本身具有无限的可延展性（Culler, 1982, p. 215）。”在后结构主义的范式中，语言被看成是一个不稳定的标示物体系，所以任何行为、记录或者意图的意义都不可能完全表达出来（Denzin & Lincoln, 1994, p. 15）。后结构主义主张，所有定性研究自反思性性质都是非客观性的、不完整的、不可避免地与研究者的解释框架、所研究的现象的情境以及用来公开研究结果的格式和风格（科研报告）密切相关。后结构主义学者的观点向那些声称具有外在权威性的文章提出了挑战。没有任何文章具有权威性或有效性，因为任何文章都可以被它内在的结构逻辑所解构。价值判断和政治，是评价研究质量的重要因素，而不是方法论上的效度。与文化、性别、意识形态、语言/文章、适用性和主张有关的权力问题是关键性的。如性别歧视和种族歧视这样的社会因素带来的影响是否在研究的时候考虑到了？研究对象是否被赋予了自由表达权？通常后结构主义研究关注的焦点是在那些不被关注的弱势群体（妇女、有色人种）。这些研究问题涉及种族、阶级和性别是怎样影响人们的生活。从一个后结构主义的角度，定性研究的效度依赖于研究

的目的（自由解放、解构）和阅读、评价研究的“解释性团体”（普通的大众、研究的参与者、科学界）。

一些作者融合后结构主义框架下对权力问题（文化、种族、性别）进行的探讨，采纳了与结构主义对立的立场，重新构建了效度这个概念（如：Lather, 1993; Scheurich, 待发表）。Lather（1993）认为，公式化仅仅是一个初步的、片面的实践，非永恒而普遍的信条。“反效度”提供了现实的多元化表征，并指出了每种表征的优势和不足之处；

“新实用主效应”则利用了动摇研究者作为知识掌握者这一信念的各种解释之间的差异；“根基性效度”，则从内部动摇权威性，非线性地呈现对现实具有不同定义的多元解释；“情境效度”假设了一个女权主义的效度，与占统治地位的、排斥女性多元化的男权主义相对立。

Woodbrook（1991，Lather在1993年曾引用）的博士论文研究就是一个后结构主义研究的例子。她研究了在高等教育中处于领导阶层的非洲裔美国女性。经过成员检查和听取同行研究报告，她对数据进行了重新审视和分析。第二版的草稿发给了研究的参与者，并且，电话采访她们的反应。Woodbrook把她作为一个白人女性研究者的观点和那些非洲裔美国女性参与者的观点并置在一起，以对话形式表达出来。她运用了Van Maanen（1988）对人种学的现实主义和批判性叙述方法，首先构造了一个现实主义的，其中强调了来自参与者的思想。然后，她又用批判性的叙述打断这个现实主义的叙述，转而强调她作为一个女权主义者的理论导向是如何塑造了她对数据的分析。最后，她用一個解构主义的陈述，其中突出了参与者对她批判性叙述的反应。其中一个结论就是女权主义意识的提升“使得女权主义运动永远成为白人中产阶级的项目，使得黑人女性和男性共有的深厚情感纽带变得肤浅（Woodbrook, 1991, p.200, lather1993年引用, p.680）。”这个后结构主义的研究阐明了在文章中听取多方面意见的方法，以及研究者和参与者之间的实时对话如何使研究者不再作为知识的唯一掌握者。

对研究范式的小结

对范式的选择是受研究者的世界观和当前感兴趣的问题影响。研究者必须清楚他们的世界观，并且选择与此相一致的定性范式。对范式的选择会影响到定性研究的策略实施以及数据分析。实证主义研究者把定性数据看成是一种信息，能够起预测和支配的作用。定量研究的一些准则（如客观性、信度、效度）被用于定性数据的收集和分析之中。

后实证主义研究者也把数据看成是能够创造或者支持与否定一个理论的信息，这个理论起到预测和支配作用。研究者采取旁观者公正的立场，这与实证主义者类似。在研究中使用与实证主义科学法相对应的方法。

解释性/建构性研究者对现实持相对主义观点，认为只存在局部和特定的现实。研究

者强调对自身和对所有参与者构念的“多重重构”（Guba & Lincoln, 1994）。研究者和参与者之间相互联系，随着研究的展开，产生研究的结果。研究的目的是为了产生一致认可的构念，比以前的更具有启发性。可信赖性这一标准被用来标志方法的严格性。除此之外，真实性这一标准被用来提高参与者的理解水平，赋予他们行动的力量。

评论性的研究者通过提升研究者和被试的意识促成双方的转变。根据Giroux（1988）分析，研究者变成了“改造过的知识分子”来排除无知，引导被试根据他们获得的知识来行动。通过公开表达价值判断，研究者和被试之间采取问答式的对话，把意义与更广阔的社会权力结构联系起来。评论性研究的质量体现它在多大程度上探讨了影响参与者日常生活的社会权力问题，以及它在多大程度上消除了无知，引导了人们的行为。

后结构主义研究者从哲学的、怀疑的立场来看待数据。研究者打破旧的观点，动摇了知识不变的概念。他们肯定了多个现实的存在，主张没有任何研究具有外部权威性。效度的概念要么就抛弃掉，要么重新定义，以便能够反映解放和解构的目的。

不同范式之间方法上的差异

这几种范式的方法存在很大的差异。合理性的问题，或者说研究如何保证其观点权威性问题，在不同的研究范式下有不同的途径。比如，三角测量法是用来加强定性研究的一个普遍方法。Denzin（1978）把三角测量法分为4类：（a）数据的三角测量（使用各种各样的数据源）；（b）研究者的三角测量（使用不同的研究方法来搜集和分析数据）；（c）理论的三角测量（从不同的角度来解释同一数据）；（d）方法的三角测量（使用多元的方法来研究同一问题）。所有三角测量的形式对实证主义和后实证主义研究者来说都是可以接受的，因为从本体论的角度，他们是分别相信现实主义和批判现实主义的。相比较而言，解释性/构建性的研究者对数据的三角测量持怀疑态度。因为他们只相信现实的相对性本质，反对用其他地方的证据来证实某个参与者的数据。Guba和Lincoln（1989）指出，“三角测量应该被看成是对反映某个事实（参与者的数量：在学校午餐项目中注册的学生数）”（p. 241）的具体数据项目的反复检查。评论性和后结构主义研究者反对使用三角测量法，因为这种方法假设了一个可以被三角测量的固定点。相反，Richardson（1994）推荐使用晶体作为定性研究过程的一个比喻：“晶体是一种棱镜，可以反射外部的东西，并在内部产生折射，产生了不同的颜色、模式、光线，投射到不同的方向（p.522）。”“晶体”画解构了传统的三角测量的思想，认为并不存在单一的真理或者固定点。后结构主义研究者揭示出一个更为极端的理解，从而对到底揭示了什么产生怀疑。

使用成员检查是另一种方法上的变式。实证主义研究者使用成员检查来确证从被试那里获得的数据具有客观精确性。后实证主义和解释性/建构性研究者使用成员检查来确证参与者的数据是否精确地表征了他们自身的事实结构（Guba & Lincoln, 1989）。这

些成员检查典型的做法就是让参与者来确证对他们的结构的描述记录是否准确。解释性/建构性研究者采用的是可靠性标准（Guba & Lincoln, 1989），他们使用了更加广泛的和交互性的成员检查，提高了参与者对调查中产生的其他结构的理解力，以及在他们不断增加的知识基础上的行为能力。评论性研究者认为，成员检查在本质上应是具有辩证性的。研究者和参与者审查综合过的观点，确证它是否更具有准确的信息。后结构主义研究者的成员检查也更加广泛。参与者除了要确证他们结构的精确性之外，还必须评论研究者对所有数据的解释。所有的这些反馈都要以文字的形式报告 [Wood brook, 1991（引自Lather, 1993）]。

定性研究的策略

研究者一旦选择了与其世界观以及想提出的经验性问题相一致的某个范式，同时也选择了如何进行调查研究的适当策略。定性研究的策略是研究者数据收集和分析的实践过程。由于篇幅所限，这里只是简要地介绍定性研究的策略。表14.1列出了与心理学研究最相关的定性研究策略以及每一种策略的侧重点。想了解更多的有关内容，可以参考Denzin & Lincoln（1994）和Patton（1990）。很重要的一点就是，尽管很多策略是和特定的范式相联系的，但一个策略可以应用于不同的范式。例如，行为调查常常运用于实证主义的范式之下；情境理论一般用于后实证主义范式；现象学、探索式、解释学和结构主义的策略典型地用于解释性/结构主义的范式之下。同样，策略常常是跟评论性和后结构主义范式紧密相连的。对策略的选择常常是基于研究的兴趣。比如，个案研究强调的是个体单元（如人）；符号交互研究考察人与人之间的交互作用；生态性研究集中在个体与环境的交互作用上；系统的行为调查方法则侧重于更大的单元，比如组织或家庭。文化的背景对民族学、民族方法学以及引导性策略尤其重要。然而，还有其他的策略也可以被用于研究文化的问题，比如解释学和结构主义。临床咨询解决的是对那些寻求医学或者心理学帮助的人们实行诊断和行为干预，可以使用很多的方法，包括情境理论和个案研究策略。总之，范式与策略之间的关系错综复杂。不过对策略的选择还是应该以范式为基础，研究者把对范式的观点融入了研究以及研究问题之中。

方法论上的共同点

各种定性范式和具体的策略千差万别，但定性研究仍然具有共同的指导方针。很多文章都提供了做定性研究最一般的方法指导（Glesne & Peshkin, 1992; Marshall & Rossman, 1995; Morse, 1993; Patton, 1990）。由于篇幅所限，这里只列出一种指导大纲，来阐述方法论上的共同点。Kvale（1995）总结出7个面谈访问的步骤，每一步都对

应效度的一个方面，他称效度为技艺的质量：

1. 确定主题：效度取决于理论原则的合理性以及理论运用于研究特定问题的逻辑。
2. 设计：研究设计和方法的适应性必须跟研究的目的相匹配。
3. 访谈：参与者报告的可信度和访谈的质量需要不断地检验。
4. 转写：数据从口头的形式转化为书面形式，对书面语言风格的选择引发了什么是有效的表述这一问题（如给那些口语中强调的语句画上底线）。
5. 解释：多大程度上提出的问题是有效的，多大程度上考验了解释的逻辑。
6. 确证：强调两个方面：确定验证研究有效性的相关形式，确定哪个团体应参与对话，讨论研究的有效性（如与参与者、科学团体、普通大众一起进行成员检查）。
7. 报告：报告的格式和内容是对研究结果的有效表征。

本章的剩余部分将着重叙述定性数据分析的一般途径。

表14.1 定性研究中选用的策略

策略	主要侧重点
个案研究	个体
人种学	人类群体的文化
民族方法学	日常生活最普遍的习俗
现象学	参与者经验的本质（内容）和结构
启发式	对现象的自我（研究者的）体验以及其他经历过同一现象的人的观点
解释学	联系具体情境，解释书面材料（文件、记录等）
结构主义	重建参与者建构的现实
情境理论 （grounded theory）	从系统收集和分析的数据中发展、完善理论
生态性的	人和环境的特异性影响了参与者所要达到的目的
符号的交互性	赋予人与人之间交互作用以意义的理解和符号
系统	一个系统（组织、家庭）作为整体的机能
行为调查	有助于一个组织和团体转型的有效行为的发展
临床调查	由诊断和治疗造成的变化，而不是大范围的社会性变化
导向性	把一些观点（如女权主义、伦理道德、黑人为中心的）运用到研究对象上，以引起一些改变

定性数据分析的具体细节

总述

尽管有一套数据分析的步骤，Patton（1990）强调这些规则并不是一成不变的：“因为定性研究的每一步都依赖于技巧、训练、灵感以及研究者的能力，定性分析最终依赖于分析者的分析智能和风格。人的因素是定性研究和分析中最强的也是最弱的方面。”

如果我们充分认识到每一个定性研究都具有独特性的话，那问题就不在于研究者是如何忠实地贯彻这些规则，而是如何完整地报告所依据的数据分析规则（Patton, 1990）。

数据分析赋予数据一定的顺序、结构和意义，这是一个非线性的即时发生的过程，通过这个过程可以找到对数据类别之间关系的一般认识（Marshall & Rossman, 1995, p.111）。数据分析发生在研究过程中的每一点上：设计研究、数据收集以及数据收集后主题和解释的精细化过程（Huberman & Miles, 1994）。

一个有用的指导方针是把研究过程分成三部分。最初的两个部分包括数据的同时搜集和分析（Marshall & Rossman, 1995, 数据搜集技巧）。研究者在对数据进行思考的同时，注意参与者的反应，改进方法上的技巧，从分析数据的同行那里获取其他的概念视角。研究过程的最后1/3是写作过程。这作为研究者的指导框架可以避免出错，防止信息的过载，并且可以很好地利用研究过程中所发现的本质特征（Lather, 1994b）。

研究对话（对研究的每一个阶段的实时讨论）在整个研究过程中自始至终都是十分重要的。研究对话包括各种各样的研究记录，如记录卡、分析性文件、磁带录音、月份报告和结构化的研究日志（Glesne & Peshkin, 1992; Patton, 1990）等。一个系统的文件体系（Glesne & Peshkin, 1992; Patton, 1990）提供了一个结构化的方案，用来组织关于研究处理的想法和定性研究的结果。使用计算机来输入这些记录、报告和日志，能够为系统地存取这些记录提供便利。反映研究多个成分的观察性、方法性、理论性以及个人的笔记能够增加研究方法上的严格性（Schatzman & Strauss, 1973）。观察笔记包括在自然观察中的一些感悟以及在与被试面谈过程中对其非言语动作所含特定信息的记录。方法性的笔记侧重于研究过程（如逻辑步骤、访谈的形式），随着研究工作的进一步展开，也可以用来改进研究方法。理论性的笔记包括在数据搜集过程中认识到的类别和主题，以及对原来研究设想支持或反对的信息。个人笔记包括研究者的信仰和猜想，它贯穿于研究的自始至终，记录研究者的观点，以避免在搜集和解释数据过程中产生偏差（Giorgi, 1970）。这个技巧常常被称为“整理研究者的思路”（Husserl, 1913/1962）。

数据分析的策略

数据分析策略有5个方面：（a）组织数据；（b）分类、确定主题和模式；（c）检验与数据不相符的假设；（d）寻求对数据的其他解释；（e）写出研究报告（Marshall & Rossman, 1995）。

组织数据

对数据进行系统组织和整理，能够让研究者：（a）获得高质量的、易于处理的数据；（b）进行分析的同时记录分析的过程；（c）研究结束后保留数据以及与数据相关

的分析 (Huberman & Miles, 1994)。所选的格式反映了数据的类别。下面列出的是数据组织的一些格式,但不是全部的格式:

1. 原始材料 (现场研究的记录、访谈的录音)。
2. 部分处理过的数据 (转写抄本、研究者思考记录)。
3. 编码过的数据 (描述少量数据的特定编码)。
4. 编码的图式 (对编码延伸性的定义)。
5. 备忘录或者其他的分析性材料 (研究者对数据概念性意义的思考)。
6. 搜寻和提取记录 (一套连接原始数据源和编码的系统)。
7. 数据表达 (以概括的方式表现信息的图表或网络)。
8. 事件分析 (对研究过程的记录)。
9. 报告文本 (陈述研究设计和方法的一系列草稿)。
10. 按时序的一般性记录 (数据收集和分析工作)。
11. 对所有先前列出的材料的目录索引 (Huberman & Miles, 1994)。

计划周密、高效的数据管理还包括在实践中注意细节,比如:(a)给录音记录贴标签;(b)准备备用的录音磁带和电池;(c)根据名字、日期和事件用不同的颜色标注;(d)在数据搜集的同时对数据进行转写;(e)对数据做至少4个计算机文件副本——1个用来备份,另外几个用于不同的数据分析;(f)同时转录和分析个案;(g)识别各个案例中相同的主题;(h)随着研究的展开,对研究的进程进行评估 (Lather, 1994b; Marshall & Rossman, 1995)。看上去不显眼的细节,其重要性往往不能低估。在数据搜集和管理中对细节的注意能减少撰写研究报告时不必要的分心。

产生类别,确定主题和模式

数据分析的核心包括确认不同的类别、一些反复出现的主题、各种想法或语言以及被试所共有的信念系统 (Marshall & Rossman, 1995)。Glesne和Peshkin (1992)建议,在产生主题时使用编码,并强调编码的递进过程:把数据组合成几个大块,在每一大块内部再分组 (建立子编码),把不同的数据列成有意义的顺序或者组成相互关联的模式。每一个主要编码的目的是为了确定一个中心概念或思想,有足够多的主要编码来涵盖所有的数据。一本按照自己的研究风格设计的编码簿有助于在整个研究过程中建立、细分和修改编码。可以将一个编码赋予任何大小的文本,从一小段到数页文字不等。确定如何将编码赋予不同的范畴,以及如何将它们联结起来,这是一件既令人激动又冗长单调的工作,需要有分析能力、创造性以及坚持不懈的毅力。

内容分析是一种常用的类别产生方法,它包括探查数据中的模式,把每种模式归类 (Patton, 1990, p.381)。研究者审查文本的每个部分,并给予标注,类似于为一本书建立一个索引。这些作为数据索引的标签和编码运用于数据相应的部分。这种索引系统

可以人工来做，对标签卡和纸版文件进行剪切和复制，或用特别为定性研究设计的电脑软件来完成。

计算机软件可以帮助研究者从事数据的管理和分析。用于Macintosh电脑上的软件包括HyperQual、Hypersoft和HyperRESEARCH；用于MS-DOS程序的软件包括AQUAD、Ethnograph、MAX、QUALPRO和Textbase α 。尽管每种程序都有其独特性，但软件程序都是专门用于描述性分析和理论构建的。主要用于描述性分析的程序（Ethnograph, HperRESEARCH, MAX, QUALPRO, Textbase α ）可以：（a）从文字处理的文件中引入文本；（b）审查每个信息的组合方式，进行编码；（c）从不同的数据源中把具有相同编码的文本分组；（d）对同一文本段采用多种编码；（e）打印出主题的各种组合，包括从一个具体的目标到若干更为广泛的目标。用来进行理论构建的程序（如，AQUAD）侧重于探索概念之间的相互关系。一旦确定了一些概念，研究者就可以使用软件来确定概念之间是否有关，是否有时间序列的模式，是否有上下属的关系，是否和3个或者以上的概念同时发生，或者是否是有因果关系。对计算机软件的详细描述见Tesch（1990）、Weitzman和Miles（1995）。

对分类、确定主题和模式有如下建议：

1. 阅读转录研究记录，寻找可能有助于修正随后数据收集工作（如访谈问题）的主题。
2. 在做编码、确定主题之前，读3遍记录，避免在细节中迷失。
3. 带一个笔记本或录音机，随时记录头脑中出现的分析想法。
4. 经常重读数据和分析文件，加上那些在头脑中涌现的想法、问题和灵感。

（Glesne & Peshkin, 1992; Lather, 1994b）

检验假设是否与数据有冲突

检验研究假设主要是指在实证主义、后实证主义以及解释性研究中对数据的分析。随着主题和模式的出现，研究者对数据进行仔细的筛选，寻找不一致的数据来挑战假设，把相符合的数据融进更大范围的构念之中（Marshall & Rossman, 1995）。对评论性和解构性的研究来说，研究者通过评估数据的质量来决定这些数据对解决研究问题有多大的帮助（Marshall & Rossman, 1995）。用数据来检验研究目的需要做到：（a）把所要研究的问题贴在研究者的工作台上，以便随时提醒主要分析的焦点；（b）比较编码簿中列出的主题和子题；（c）使用简单的矩阵来总结不同来源的数据的主题（Glesne & Peshkin, 1992）。贴在工作台墙上的纸可以用来记录不同的数据来源、所出现的主题以及相容或不相容的模式，在分析的过程中建立可能的模型。

寻求对数据的其他解释

为了提高分析的广度，控制研究者的偏差，在编码时应该考虑采纳多种角度，包括：（a）从不同的理论角度编码数据；（b）让多个研究者——包括不同背景的人（如性别、种族）来编码相同的数据；（c）要求被试对编码提供反馈，同时给予一些建议（Glesne & Peshkin, 1992; Lather, 1994b; Marshall & Rossman, 1995）。从不同的角度来处理数据、记录多个成分可以提高研究的可信度。

一旦从数据中产生了一套主题，从余下的文本中寻找不相容的信息就很重要。比如，如果6个研究对象中有4个都强调他们的早期学校经验是正性的，研究者就应该做一个反面的个案分析，在其他两个研究对象中寻找相反的证据。在研究报告主题应该考虑到所有的个案，采用恰当的陈述（如，6个参与者中的4个报告了正性的早期教育经验）。有必要注意那些与总体不太相协调的数据，这样可以避免做过于简单化的解释，简单化的解释会掩盖主题的复杂性；有必要让读者自己对数据的解释进行判断，从而最大限度地提高研究的可信度（Glesne & Peshkin, 1992; Lather, 1994b; Marshall & Rossman, 1995）。

撰写研究报告

研究者从数据收集到构思文章（研究报告），至少要考虑3个问题：文章的组织、文章的陈述以及文章的合理化（Denzin, 1994）。

首先，文章的组织必须考虑到“要报告什么”和“怎样报告”两个问题，信息的来源包括：从被试那里搜集的数据；研究者的理论、现场和编码记录；任何通过研究成员检查所得到的信息以及后续访谈的信息。

其次，文章的陈述解决的是文章的声音和文章读者的问题。被试的声音，经常被称之为“他者”，和作者在上下文中的角色都是要特别考虑的问题。正如Denzin（1994）所提到的那样，文章的陈述总是研究者的自我陈述，因为正是研究者决定了所要陈述的内容和顺序。建立来自多个角度的观点（而不是单方面的）可以部分地解决这个问题[如Lather（1993）引用的Woodbrook的后结构主义研究（1991）]。不过，表述的问题不可能完全解决，因为有两个不可调和的方面：一方面是将文章的效度看作“同构性和真实性”的一种形式；另一方面，则是一个铁的事实，即所有的文章都具有历史的、社会的、政治的以及文化的限制。（Lincoln & Denzin, 1994, p. 582）。

再次，合理化是指文章与公认的标准相对应的维度，这些标准包括认识论上的效度或被科学团体所接受的有效范例。先前已提到过，每种范式都有各自合理化的标准。Lincoln和Denzin（1994）提出了与合理化有关的几个问题：“文章是否忠实于它所要表述的情境和个体？文章是否能做到不囿于研究者的个人兴趣，而且包含被研究对象的兴趣

(p.578)?”研究者在把他们的数据和解释变换到文章里去的时候,必须要考虑到这3个问题。

针对不同的研究目的,应当选择不同的文章组织形式和书写风格。Hammersley和Atkinson(1983)推荐使用以下的方法来组织文章:(a)自然历史顺序(对实地调查过程很有益);(b)按时间排列事件(当时间进度对研究目标很重要的时候十分有用);(c)变换焦点(从数据的描述到理论的抽象);(d)把叙述和分析分离开来(用不同的段落来陈述研究过程和研究理论);(e)按主线或者主题组织(对分析数据得到的每个主题进行深度讨论)。不同的写作风格方法被称为记述,包括现实主义的(权威的观察角度)、独白的(解释、强调研究者对数据的选择性观点)、印象的(创造性、艺术地回忆现场调查体验)、评论的(对研究所基于的社会、政治和经济问题进行阐述)和正式的(对理论清晰地建构、检验或者展示)(Van Maanen, 1988)。

总 结

单是懂得数据分析过程对定性的数据分析来说是远远不够的。定性研究过程的各个阶段都是相互联系的,并且与构成研究框架的范式紧密相连。因此,定性研究者必须站在宏观的角度看问题,因为范式决定了研究的设计、策略、数据分析和对结果的报告。本章回顾了对定性研究中5种研究范式,强调了方法论问题以及与方法论相关的策略问题。另外,也指出了收集和分析定性数据的一般步骤。

定性研究能让研究者达成实用和准确的结合。定性范式提供了研究现实的社会性、构建性的系统方法,注意到了个体的变异性和复杂性,提供了获得并验证知识与理论的发式途径。因此,定性研究是一种对世界进行系统、科学研究的切实可行的途径。

推 荐 阅 读

要了解关于本章主题更详细的内容,参见:Denzin & Lincoln(1994), Glesne & Peshkin(1992), Huberman & Miles(1994), Hill(1994), Van Maanen(1988), 以及Weitzman & Miles(1995)。

参 考 文 献

- Cook, T., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin.
- Culler, J. (1982). *On deconstruction*. Ithaca, NY: Cornell University Press.
- Denzin, N. K. (1978). *The research act: A theoretical introduction to sociological methods*. New York: McGraw-Hill.
- Denzin, N. K. (1989). *Interpretive interactionism*. Newbury Park, CA: Sage.
- Denzin, N. K. (1994). The art and politics of interpretation. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of*

- qualitative research (pp. 500-515). Thousand Oaks, CA: Sage.
- Denzin, N. K., & Lincoln, Y. S. (1994). Introduction: Entering the field of qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 1-17). Thousand Oaks, CA: Sage.
- Geertz, C. (1973). *The interpretation of culture*. New York: Basic Books.
- Gibson, R. (1986). *Critical theory and education*. London: Hodder & Stoughton.
- Giorgi, A. (1970). *Psychology as a human science: A phenomenologically based approach*. New York: Harper & Row.
- Giroux, H. (1988). *Schooling and the struggle for public life: Critical pedagogy in the modern age*. Minneapolis: University of Minnesota Press.
- Glesne, C., & Peshkin, A. (1992). *Becoming qualitative researchers: An introduction*. White Plains, NY: Longman.
- Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Communication and Technology Journal*, 29, 75-92.
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.
- Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105-117). Thousand Oaks, CA: Sage.
- Habermas, J. (1971). *Knowledge and human interests* (J. Shapiro, Trans.). Boston: Beacon.
- Hammersley, M., & Atkinson, P. (1983). *Ethnography: Principles in practice*. New York: Tavistock.
- Hill, C. E. (Ed.). (1994). Special section: Qualitative research in counseling process and outcome. *Journal of Counseling Psychology*, 41, 427-512.
- Huberman, A. M., & Miles, M. B. (1994). Data management and analysis methods. In N. K. Denzin & Y. S. Lincoln (Eds.), *Qualitative research handbook* (pp. 428-444). Thousand Oaks, CA: Sage.
- Husserl, E. (1962). *Ideas*. New York: Colliers. (Original work published 1913)
- Janesick, V. J. (1994). The dance of qualitative research design: Metaphor, methodology, and meaning. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 209-219). Thousand Oaks, CA: Sage.
- Kvale, S. (1995). Validation as communication and action. *Qualitative Inquiry*, 1, 1-19.
- Lather, P. (1992). Critical frames in educational research: Feminist and post-structural perspectives. *Theory Into Practice*, 31(2), 87-99.
- Lather, P. (1993). Fertile obsession: Validity after poststructuralism. *Sociological Quarterly*, 34, 673-693.
- Lather, P. (1994a). Critical inquiry in qualitative research: Feminist and poststructural perspectives: Science "after truth." In B. F. Scrabble, W. L. Miller, R. B. Addison, V. J. Gilchrist, & A. J. Kuzel (Eds.), *Exploring collaborative research in primary care* (pp. 103-114). Thousand Oaks, CA: Sage.
- Lather, P. (1994b). *Introduction to qualitative research in education*. Class lecture, The Ohio State University, Columbus.
- Lincoln, Y. S., & Denzin, N. K. (1994). The fifth movement. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 575-586). Thousand Oaks, CA: Sage.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.
- Morse, J. M. (1993). *Critical issues in qualitative research methods*. Thousand Oaks, CA: Sage.
- Marshall, C., & Rossman, G. B. (1995). *Designing qualitative research* (2nd ed.). Thousand Oaks, CA: Sage.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Newbury Park, CA: Sage.
- Richardson, L. (1994). Writing: A method of inquiry. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 516-529). Thousand Oaks, CA: Sage.
- Rose, J., & Jevne, R. F. J. (1993). Psychosocial processes associated with athletic injuries. *Sport Psychologist*, 7, 309-328.
- Schatzman, L., & Strauss, A. (1973). *Field research*. Englewood Cliffs, NJ: Prentice Hall.
- Scheurich, J. (in press). The masks of validity: A poststructural interrogation. *International Journal of Qualitative Studies and Education*, 9(1).
- Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage.
- Strauss, A., & Corbin, J. (1994). Grounded theory methodology. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 273-285). Thousand Oaks, CA: Sage.
- Tesch, R. (1990). *Qualitative research: Analysis types and software tools*. Philadelphia: Falmer.
- Van Maanen, J. (1988). *Tales of the field: On writing ethnography*. Chicago: University of Chicago Press.
- Weitzman, E. A., & Miles, M. B. (1995). *Computer programs for qualitative data analysis: A software sourcebook*. Thousand Oaks, CA: Sage.

第 15 章

统计研究和探索的基本指南

——统计分析和选择

Robert A. Yaffee

在设计和进行研究的时候，选择适当的统计分析方法，取决于所考察的问题和其中所含变量的性质。不论研究是验证性的还是探索性的，研究者都应该在收集数据之前努力做好他们的设计方案。在计划任何一个项目时，都需要作出一系列的选择。对其中的许多选择来说，统计检验可以提供一定的参考信息。在设计一个实验研究时，第一，研究者应该估计预实验的样本大小、调查（或实验）样本大小、样本的设计和权重，以及对每一个理性假设的统计检验方法。第二，预研究可以考察应答率和覆盖率，并评估实验或调查中用到的单个项目和一组项目的效用。其中应答率和覆盖率是在设计一般样本大小时必须考虑的。第三，一旦收集到数据并进行了输入，那么就需要进行包括误差检查和纠正在内的数据整理工作（参考 Dollinger & Dilalla，本书第13章）。如果存在缺失数据，就必须建立一个替代的标准并依此执行。最后，对假设进行检验。对学习心理学的大学本科生和研究生来说，本章只能作为统计学的基本指南。

本章的主要内容是如何选择能够解决研究问题的统计检验。我们将分析工作分成3个部分：数据收集前的工作、假设检验前的工作以及假设检验的目的分析。如果可能的话，我会谈谈诸如SAS（SAS Institute, Inc., 1990）、SPSS和SYSTAT（SPSS Inc., 1989, 1990）、BMDP（Dixon, 1990）、LIMDEP（Greene, 1992）之类的统计软件。本章余下的章节结构如下：在简要介绍数据收集前的分析之后，我将介绍假设评估前需要完成的初步分析。本章的主要内容是假设评估检验的选择问题。在这一节里，我将回顾一下对单个因变量的检验。这些检验既包括对单个自变量的检验，也包括了对多个自变量的检验。然后介绍多个因变量的检验。此外，在模型评估前对假设进行检验也是很好的练习。我认为，对不同统计模型要进行适当的诊断和评估。当违背参数假定时，我们就要用非参数分析代替。对大多数分析而言，两种方法都会介绍。但是在使用非参数检验时，应该首先评估检验的相对效力，以保证假设和效力之间的平衡是可以接受的。

收集数据前的分析处理

在收集数据之前，很重要的一点是对所需要的样本大小、取样设计和样本权重进行多方考虑。这里讨论样本大小和统计效力问题，McCready（第8章）讨论了关于采样设计和样本比重问题。

对样本大小的要求

为了估计对样本大小的要求，研究者应进行一个效力分析。一个统计检验的效力是指它拒绝一个错误的虚无假设的能力。Cohen（1988，pp.14-16）定义了统计推论的4个相关因素：效力（power）、显著水平（ α ）、样本大小（ N ）、效应测度（ES）。效应测度是指一个心理学现象存在的程度。如果其他条件不变，样本越大，效应测度就越大，统计检验就越有力。而统计检验的效力越大，能够被发现的效应测度就越小。效力越大，检验导致错误接受一个虚无假设的可能性就越小。所以，一个研究者要得到足够的统计效力，就需要足够大的样本（经验性的拇指定律建议效力不低于0.80；Cohen，1988）。4个因素中当其中3个因素被确定时，第4个也就确定了。

较简单的效力/样本大小的计算可以笔算。假设几个研究者正在分析比例，并用百分比的形式报告。他们可能决定可接受的错误量不超过估计总量的3%。被评估的样本大小（ N ）是否达到这个错误是可以通过计算比例的置信区间公式得到。对统计效力更详细的估计，推荐用表格或电脑程序。Cohen（1988）提供的表格是最全面的；Borenstein和Cohen（1988）、Borenstein、Cohen、Rothstein、Pollack和Kane（1992）、Lenth（1987）提供了这样的程序，而“科学软件”公司提供了一个更先进的程序——STATPOWER（Bavry，1991）。Godstein（1989）对估计样本大小或效力的几个PC程序进行了比较，可参考他的工作从而得到更进一步的信息。

虽然效力分析决定了必要的样本大小，但是对预研究和一般调查（实验）所需的接触样本大小要分别估计。预研究可以告诉我们应答率和覆盖率，以便研究者在所需样本大小上有所增加，从而获得接触样本的大小。用来访谈和观察的样本应该有必要的大小和效力（Deming，1960；Kish，1965）。

假设评估前的分析处理

收集到数据之后，在假设检验之前，研究者需要完成整理数据、单项目分析、量表的建构和评估。而且，应该有适当的对缺失数据进行筛选和替代的策略。在本书第13章中，Dollinger和DiLalla更详细地讨论过其中一些初步分析。本节中，我将叙述单项目分

析和量表评价。

单变量项目分析

在多项目的量表构成之前，需要评估每一个项目。通常，如果一个项目有超过30%的数据缺失，就应该从研究工具中删除。另外，替代缺失数据也可能是合适的。如果这个项目的分布是近似正态的，在代替缺失数据时，可使用平均值。也可以利用数据库中其他变量，通过回归的方法来获得这个变量的缺失值（例如，利用BMDP子程序AM）。进行这样的估计要用到一个回归方程（Anderson, Basilevsky, & Hum, 1983）。如果分布不是正态的，例如一个有极端值的项目，则可以通过删除两侧的极端值而后计算出一个“调整”的平均值。如果分布偏斜于某一侧，那么可以用中数来代替缺失数据。

量表分析：信度和效度

研究者经常把项目集合成量表来分析数据和检验假设。对单项目进行分析之后，研究者利用各种相关关系来估计调查或实验中的测量结果。从大量可利用的相关系数中选出适合研究兴趣的系数是极其重要的。

两两变量相关是计算变量之间相关性的一种方法。相关系数的选择取决于变量的具体特点。如果两个变量都是连续的，或者至少是等距的，研究者可以画出这两个变量的联合分布图来验证两者关系是否是线性的。如果呈现的是非线性关系，研究者就要检验不同的函数形式来拟合它们的关系，或者根据最优的函数来变换其中一个或两个变量，使它们之间的关系线性化，当然还可以使用非线性回归方法。基本的系数是Pearson瞬间产物系数（PPMC），它可以用来衡量当一个变量发生变化时，另一个变量随之改变的程度。一个常用的变换是把Fisher的 r 转换成 z ，这个变换可以使样本的相关系数分布正态化。对PPMC的显著性检验是 t 检验，它可以用来检验一个相关系数与零有没有显著差异。

如果两个变量不是等距的，则存在几种可能的情况。第一种情况是在相同的水平上测量两个变量。如果两个变量都是有序的，变量间没有明显的自变量和因变量的关系，且研究者拥有的样本总量比较小（如， n 小于30），来估计相关关系的大小，这个大小定义为单调性——两个变量都增加或者都减少。一些非参数相关性系数包括：Spearman的 ρ ，Kendall的 τ_a 或者 τ_b ，Stuart的 τ_c ，对称性 λ ，Goodman和Kruskal的 Γ ，polychoric correlation。其中只有 τ_a 和 Γ 完全相等这种情况两个变量都没有发生变化的情况。一般来说，对相同的数据， Γ 的数值最大， τ_a 值最小。 τ_b 纠正捆绑效应， λ 和 τ_c 可以处理具有不同数量的数值的两个变量之间的相关性。对变量是分类变量的情况，polychoric correlation可以估计出两个正态分布的变量间的关系（Jöreskog & Sörbom, 1988）。第二种情况是，如果存在明显的自变量和因变量，研究者就必须弄清楚哪一个是自变量，哪一个

是因变量。对这样的不对称关系，可以用Somer的 D 或者不对称 λ 来估计两个变量间的关系。

当两个变量都是称名/分类变量时，可以用Goodman和Kruskal的 λ 、Cramer的 V ，或者相依系数。Cramer的 V 的值在0、1间变化，而相依系数一般在0.87处达到它的最大值。 λ 的值也在0、1间变化，它表示当已知自变量的数值时，预测一个因变量时的误差减少的比例。我们可以认为，对称 λ 是两不对称形式的平均值。如果称名变量是二分的或者二元的，那么可以用 ϕ 值、四分法的、或者Yule的 q 值来估计相关的程度。当PPMC系数应用到二进制数据时，它就简化成了 ϕ 系数。在很多跟踪一群人的群组研究（cohort study）中，可以用相对风险来估计事件收敛率的比例；在case-control研究中，相对风险可以表示某一个个案事件与控制事件之间的比率。也可以用一个连续调整的 χ^2 系数来检验显著水平。对一个不平衡、倾斜的或者稀疏的样本，Fisher的检验将得到精确的显著水平。

当两个变量的测量水平不同时，相关系数的选择变得更加复杂，这是由于相关系数的选择依赖于变量的不同测量水平以及变量之间关系的对称性。分析者们通常使用用来处理低测量水平而不是高测量水平的相关系数。例如，两个变量中的一个可能是二分的，而另一个是等距、比率、顺序型或者称名的。假设另一个变量在自然界中是连续的，如果调查者能够证明二分变量是以正态分布为基础的，那么二列相关系数是合适的。另一个系数是点二列系数，条件是二分变量代表了真实的二分（Andrews, Klein, Davidson, O'Malley, & Rodgers, 1981; Joreskog & Sorbom, 1988）。如果一个变量是二分的而另一个是顺序的，那么 η^2 、 ϵ^2 或者Hay的 ω^2 是适当的。如果一个变量是二分的而另一个是称名的，那么相依系数、Cramer的 V 、 τ_c 或者不对称 λ 系数都是合适的。

进一步讲，并不是只有二分变量可以与其他测量水平搭配。需要计算相关的两个变量有可能一个是顺序变量，而另一个是连续变量或一个称名变量。如果一个变量是顺序型的而另一个是连续的，那么多系列相关系数是一个合适的系数。当然，使用的顺序变量需要假定是符合正态分布的（Andrews et al., 1981; Joreskog & Sorbom, 1988）。此外，Spearman的 ρ 、Kendall的 τ_a 或 τ_b 、Stuart的 τ_c 或Goodman和Kruskal的 Γ 系数也是适当的。如果第二变量是称名的，那么可以使用 λ 、 η^2 、 ω^2 或者 ϵ^2 （Kirk, 1994; Winer, Brown, & Michels, 1991）。如果两个变量中一个是顺序的而另一个在自然界中是连续的或者是整数型的，那么 η^2 系数是合适的。

量表评估

量表由项目组合而成，每个项目衡量心理结构的一部分。一个量表可以用不同的方法来分析，以判断这些项目是否组成了一个或是多个量表。但是，开始总是应该先估计并报告量表的信度。信度在这里的定义是实际分数方差相对于观察到的方差的比率。用系数来衡量的话，它的值在0、1之间变化。有三种基本的信度：稳定性、同等性和内在

一致性。测量项目在时间上的稳定性可以由重复测验的相关性来评估。可以用同类相关来估计多个评分者之间或重复观察的一致性。可以用Cohen的 κ 来评估两个或更多评分者的一致性，这种方法对偶然一致性也进行了控制（对两个以上的评分者，必须使用多位评分者（ κ 系数））。同等性的系数，通常是PPMC，与某单个范畴（如工作满意度）中的两个子项目间的关系有关。与某单组项目有关的内在一致性系数包括分半估计、Cronbach的 α （ α ）、Kuder-Richardson KR-20/21以及Cochran的 Q （如果项目是二分的）等。

Nunnally在早期的研究中认为，0.50、0.60或更大的信度对一个量表是足够的，但是后来他又称应该在0.70以上（Nunnally & Bernstein, 1993）。当相关性的假设不太可能满足时，可以用Kendall的成对等级系数，Kendall的一致系数或者Kendall的协调系数（Siegel & Castellan, 1988）。SPSS软件中有一个计算信度的程序可以计算其中的一些系数，SAS软件在它的相关性程序中有一个Cronbach's α 的选项。

可以用因素分析来对量表进行评估（Rummel, 1970）。另一个选择是采用多维量表法，这是一种通过变量间的相似性抽取维度的方法。通过因素分析，可以得到因素的数量、性质以及因素间的相关度。每一个因素组成了量表的基础，每个被试在每个因素上的值都可以当做测量结果来使用。因素的载荷反映了潜在的结构与每一个可观察的标志变量间的关系。因素结构，即变量之间的相互关联的模式，构成了项目间内在一致性的基础。它代表了高度相关的变量的均一性。维度或提取的因素为理论构念的结构效度提供了证据（Cook & Campbell, 1979）。如果一个结构的成分维度无法在因素结构中找到，就表明该结构缺乏效度。因素分析提取的因素可以支持也可以不支持假设的维度，这种结果也是对结构效度的检验。因素权重可用来代表每个变量在一个因素上的位置。如果一些项目用一种方法测量而另外一些项目用别的方法测量，那么，利用多特质—多方法相关矩阵就可以对收敛和判别有效性进行检验。

量表建构的另一个方面是单个项目的外部相关。一个项目的外部相关，表示该项目与一个外部标准的相关程度，以及项目到什么程度就可以推广到其他人、人群、环境、时间上（Cook & Campbell, 1979; Pedhazur & Schmelkin, 1991）。如果从不同样本中抽取相同的因素结构，我们就可以认为量表的结构具有外部效度。研究者往往从样本数据中保留一些数据，用来对因素分析得到的因素结构进行交叉检验。

检 验 假 设

当这些初步分析完成后，就可以对研究的理论问题进行检验。本书在第16章由Dickter和Roznowski、第17章由Levy和Steelman更详细地介绍了如何使用基本的和高级的统计分析。在这一节，我将介绍单个因变量的单样本检验、成对样本检验、双样本检验以及一般的线性模型，包括方差分析（ANOVA）、回归法、协方差分析

(ANCOVA)、模型建立(model building)和多变量分析(multivariate analyses)。

单样本检验

在一些情况下,研究者希望能确定一个样本是否来自于带有一定参数(如百分比或平均值)的人群。也有些时候,研究者希望知道他/她离散的或连续的样本是否来自于一个正态分布的人群。换句话说,研究者都假定得到的样本具有潜在的连续的正态分布。可以使用Shapiro-Wilks W或者Kolmogorov-Smirnov统计来检验分布是否同理论上的正态分布存在显著的差异。而且,可以使用一些图形来估计分布的图形形状,特别是标有正态曲线的柱状图(SPSS频率计算)、盒状图和须状图(或一般的概率分布图,SAS单变量;SPSS MANOVA或者回归程序中的残差分析)。

一些测量的变量是二元的或二分的,仅仅只有两个数值。研究者想知道获得一个或另一个结果的概率。如果一个人摇骰子,他就可以知道摇5次中得到两次3点的概率。选择的检验是二项式检验,它可以给出得到这个结果的概率。利用对每一个少于两个3点的可能性的检验,可以给出得到5次中小于等于两个3点结果的概率。当样本量足够大时,连续性的就可以进行修正。

当变量是分类变量时,研究者就希望能确认这个结果是否与他预期的结果存在显著差异,那么,他可以用卡方(χ^2)检验,检验观察到的频率与预期频率之间的差异。

在实验设计中,有时研究者要考察控制组与实验组之间分配的随机性。可以用一个单样本运行检验,来达到这个目的。对一个群体连续分配的运行时间越长,分布就越不可能是随机的。列出的数值给出了不同运行的长度临界水平。对大于20的样本这个检验是适用的。

成对的单样本检验

在一些情况下,一个变量被重复测量。研究者就希望能确定在某个影响或干涉后变量值是否会发生改变。通常,在实验干涉前后会重复使用同一群人。有时,会使用一种不是最佳的方法,即两组人之间在可能影响实验结果的无关变量上做匹配。无论哪种情况,当样本中检测的变量是连续的(等距的或等比率的)时,如果我们需要检查平均值在前测与后测之间变化,配对 t 检验是最有力的检验方法。检验结果的有效性建立在特定的假设基础之上。这些假设包括样本的随机性、正态性和方差齐性。如果总体不是正态分布的,但是样本足够大,那么平均值的采样分布就变成正态的了。大约40的样本量就已经足够大了(Hays, 1973)。

如果样本太小而且不是正态分布的,那么就应该选择一个非参数检验。下面按效力—效率递减的顺序讨论6种非参数检验。效力—效率是指最有效力的统计检验的事件数目除以可能达到相同效力的统计检验的事件数目。有6种最常用的非参数检验。最好的

是Wilcoxon 符号等级检验，这个检验同 t 检验相比较，有95.5%的效力—效率。其他的检验效力较小，它们是McNemar、符号、二项式、Cochran的 Q 、Kendall的和谐系数和Friedman的双尾检验。

6种方法具有不同的特点。Wilcoxon检验容易计算。把第一次和第二次重复测量之间的差值排列起来。正的等级记上正号，负的等级记上负号。正的总和的差除以标准差得到 z 值。如果 z 的绝对值超过了1.96，那么结果就是显著的，拒绝无效假设。对相同等级做一个修正后，这个检验就是一个不受样本分布制约的检验方法，可替代 t 检验。为了检验称名变量或顺序变量变化的统计显著性，可以用不受分布限制的McNemar检验。这个检验可以得到在预测和后测之间实际变化值与预期变化值的 χ^2 。它经常还要对连续性进行修正，因此样本量必须足够大，使得没有太多的单元其预期发生次数少于5次。这个检验的效力—效率在63%~95%之间变化，具体数值与样本大小有关。其他两种非参数检验是符号检验和二项式检验。当样本大小是6时，这两种检验的效力大约是95%，随着样本大小的增加，逐渐减小到63%左右。Cochran的 Q 检验是一个 k 样本的比例检验，可以用来确定由一个称名变量或顺序变量内分布决定的比例是否会随着时间而发生显著变化。这个检验是一种修正的 k 列 n 行的 χ^2 系检验，也可以用来检验二分变量。Kendall的一致系数 (u) 对两个判断或成对的观察是有效的。Kendall的协调系数 (W) 数值在0、1之间，是一个替代完全组内重复测量设计的非参数检验方法。

双变量方向模型

除了上面讨论的双变量联合模型之外，很多情况下因变量与一个自变量之间的关系决定了这个预测变量是否应该被包含一个更复杂的、具有多个自变量的模型之中。在因变量与理论上可能相关的自变量之间进行一系列两两变量检验时，又被称作“变量筛选”。只有显著的相关关系才会在模型中被使用。在一个遗留样本或新样本上进行交叉验证是一个很明确的要求。

如果考虑使用的模型是一个有向模型，首先要考虑的问题是因变量和自变量各自的测量水平。假设因变量是连续的，那么这个问题就变成自变量的测量水平了。自变量可能是连续的、离散的或者类别的。假设自变量取两个值，比如有或没有实验处理、干涉，统计效力高的参数，独立样本 t 检验就应优先采用。这个检验的前提假设是，当自变量在其正常数值之类时因变量是独立的、正态分布的。如果方差是齐性的，那么就可以用 t 检验；如果方差是不齐性的，那么就要对自由度进行修正，来获得各自独立的方差。为了保证检验结果的有效性，样本应该是正态分布的。但是，如果样本大小足够大而且方差相等，那么，这个检验允许适度违背正态性。

如果根本违背了参数性假设，那么，就应该换用一个不需考虑分布的非参数检验；如果因变量是序列型的或者能作为序列型的来处理，同时自变量是二分的，那么可以使

用Mann-Whitney U检验。它还可对相等等级进行修正。一般来说,非参数检验的唯一前提假设是观察间的相互独立(Siegel & Castellan, 1988)。

一般线性模型

一般线性模型包括方差分析(ANOVA)、回归和协方差分析(ANCOVA)。如果因变量是连续的,自变量有3个或更多的水平,那么可以用单向方差分析。如果存在多个类别自变量,那么就要用因素方差分析来分析主效应和交互作用。为了保证分析结果解释的正确性, F 检验的假设必须满足。抽样通常被假定是随机的,除非研究是那种对被试的选择设定了标准的临床性质研究。残差被假定是正态分布的,这可以用Shapiro-Wilks检验(对较小的样本)或Kolmogorov-Smirnov检验(对较大的样本),考察样本的正态性。由于这个模型能允许适度违背正态假设,所以可以用一个叠加了正态曲线的柱状图或一个正态概率图形来评估正态性。第三个比较重要的假设是,残差的方差应该相等。残差的同质性可以用Hartley的 F 检验(大方差除以小方差)或者对非正态不太敏感的Levene检验(Kirk, 1994; Snedecor & Cochran, 1980)来考察。残差可以用标准残差与因变量(和/或自变量)的相关图来考察。如果分析中的平均值受到极端值的影响而失真,可以在个案图形、Mahalanobis或Cook距离的协助下寻找有问题的个体。如果条件允许,可以将极端值从分析中删除。BMDP的程序7D可以对修正后的平均值进行方差分析。最后,当假设不满足时,可以利用因变量的转换来使得残差的分布更加贴近正态或均一(如布洛姆转换余弦)。

如果前提假设得到满足,总体检验是显著的,那么,研究者必须决定自变量的不同水平之间各组因变量的均值中哪一组是显著不同的。可以采取两种方法。一种方法是利用先验或计划的对比来检验平均差异。对复杂的正交对比,有多重 t 检验、多重 F 检验、robust Behrens-Fisher和Welch检验;对非正交复杂对比,Bonferroni和Dunn-Sidak检验是适用的;对成对比较,可以使用水平间的Dunnnett检验。第二种方法是利用非正交多重比较来检验平均差异。当使用这种检验时,研究者有必要根据进行成对比较的次数来计算实验的误差率,然后他或她应该用Bonferroni或Dunn-Sidak修正来决定正确的对比显著水平。对复杂的比较,一个通常使用的保守检验就是Scheffe S 检验。这个检验要求方差的齐性,而且往往只有较小的效力。也可以使用较有效力的Brown-Forsythe (BF) 检验。这两个检验不需要相同的样本大小。作成对比较,有Fisher最小显著差异检验(LSD)、Turkey真实差异检验(HSD)(检验最大和最小平均值之间的差异)、Spjotvoll和Stoline的 T 检验、Tukey-Kramer修正真实差异检验(HSD)(后两个检验在样本数不同但满足方差齐性假设时被使用)。当样本数不同且方差也不满足齐性假设时,可以考虑使用由Games-Howell/Tamhane从Behrens-Fisher统计中发展出来的程序,其自由度在这个程序中

经过了Welch修正。逐步的或分层的显著性检验应该使用Newman-Keuls 检验 (Kirk, 1994)。在主要的统计程序软件包中可以找到许多这些检验的子程序。

如果数据明显违背前提假设,那么研究者可以使用非参数检验。如果因变量是包含两个或多个水平的顺序变量,Kruskal Wallis单因素非参数顺序方差分析是一个合适的选择,它可以检验自变量不同水平的等级平均值之间的差异显著性。这个检验相对于 F 检验有95.5%的效力—效率 (Siegel & Castellan, 1988)。如果这些等级总和之间存在显著差异,当相等等级被纠正后,它就表明存在显著差异。

如果分析者发现一个双因素方差分析、单因素随机分组设计或者分组方差分析由于违背了潜在假设 (特别是重复测量中的复合对称)或被试间和被试内效应的误差定义使用不当,那么,他或她可以用Frideman双向序列方差分析。这些检验适用于两个顺序变量或一个顺序变量和一个类别自变量的情况。

回 归

如果因变量和自变量都是连续的,那么比较合适的检验是简单 (双变量) 回归。回归也需要特定的前提假设,包括随机采样、独立观测、线性、误差的非相关性、残差的正态性、方差的同质性以及自变量之间没有多重共线性。这些假设都可以用前面提到的方法来检验。如果假设条件不满足,那么,研究者就可以用非参数回归分析 (Hollander & Wolfe, 1973; Hettmansperger, 1984); 如果只有两个自变量,分析者就可以用Kendall偏相关;如果有两个或两个以上的自变量,而且它们是不相关的,就可以选择Brown-Mood方法,以此来发现基于每一个 x 轴和相应 y 轴上以中数分半的数据回归线的斜率和截距 (Daniel, 1990), 并选择Theil斜率估计和不考虑分布的检验来考察其显著性 (Hollander & Wolfe, 1973)。在假设条件不满足时,这些方法都是参数检验的有用补充。

建模和回归

当要揭示的关系是一个因变量同几个自变量之间的关系,分析就依赖于变量的测量水平。偏相关揭示的是,当第三个变量保持恒常时,两个连续变量之间的关系。在有一个等距因变量和几个连续自变量的情况下,多重相关系数表示了关系的大小。由于在模型中加入新的自变量会增大这个系数,对多重相关系数进行调节或缩小能补偿自由度的变化。多重相关平方 (SMC) 系数被称作“决定系数”或 R^2 ,它表示因变量的变异在多大程度上能被模型中的自变量解释。一个偏相关系数的平方反映了增加一组新的变量而引起的 R^2 变化。变更的 F 检验可以用来作为这些综合统计的显著性检验。通过添加能显著增大决定系数的变量,就可以得到普通最小平方法 (OLS) 和多维线性回归模型。

可以用4种计算机算法来为OLS模型选择变量。具体说来, 计算机统计软件提供了向前、向后、逐步、强行进入的方法来为模型筛选变量。一方面, 逐步法会错误地漏掉一些正确说明和解释所必需的成分; 另一方面, 强行进入方法对层次模型、理论建立以及假设检验都是有用的。

在多重线性回归模型的前提假设得不到满足的情况下, 仍然有一些其他替代的方法。如果违背了线性假设, 那么可以用某种函数变换来线性化因变量与各个自变量之间的关系; 如果变换还不能建立线性关系, 那么就用多重称名或者非线性回归, 不同的效力表示特定的非线性形式。当确实存在多重共线性时, 就用工具变量代替相关联的自变量。对高度相关的自变量进行主成分分析能够得到在回归分析中可以用作为自变量的成分值。在同质性和正态假设不满足的情况下, 可以做方差的稳定化和正态化变换。如果方差仍然保持非同质性, 那么可以用加权的最小平方作为估计方法。经过Durbin-Watson检验、Durbin的 m 检验或自相关诊断的相关残差, 可能需要优先差分或者其他自回归方法。但是, 如果这些方法都不能排除这一问题, 那么可以通过一系列的PPMC来序列性地考察因变量和每一个自变量之间的关系。

由一个包括所有有关经验参数的一般模型开始, 连续检验和删除非显著的效应, 直到获得一个最简洁的模型, 这个建模的过程能够最大程度地扩展理论的范围和效力 (Hendry & Richard, 1982)。虽然也有例外, 但模型修正通常能提高效力和排除非显著效果。当需要有主效应 (main effects) 来确定交互作用的时候, 不能以交互效应对主效应的显著性造成的抑制为由, 排除成分主效应。成分主效应对确定相互作用来说是必须的。例如, 较低水平的相互作用对确定较高水平的相互作用是必须的 (Cohen, 1978)。否则, 非显著的效应会从模型里被除去。

利用回归程序或协方差结构分析 (比如利用EQS、CALIS、RAMONA或LISREL), 可以发展出路径分析模型。双变量关系可能掩盖先行或中介变量。路径模型将一个因变量的总效果分解为直接的、间接的和虚假效应。它们能够显示一些变量如何在一个先行变量和一个结果变量之间起到中介作用。路径模型由一组OLS回归方程组成, 但现在通过LISREL, 也可以对顺序或称名变量建立模型。这些分析对发展和评估假设模型是有用的。

其他问题可能影响着这些模型。Leamer (1983) 指出, 错误可以导致模型的虚弱。在这样的模型里, 一个变量的加入可能会导致其他变量的回归系数发生的剧烈变化。Leamer建议, 每一次只关注一个重要的变量, 用不同的组合方式增加次要变量, 记录下被关注变量的回归系数大小的变化。最小和最大数量就构成了极值边界。如果整个模型中边界的范围超过了变量的置信区间, 那么, 这个模型就是虚弱的、不可靠的。虚弱成分是可以从模型中排除的。用这种方法可以提高模型的强度, 也可以增加解释的说服力。

当存在一个因变量和多个自变量 (IVs) 时, 因变量的测量水平可以是二分的、顺序的或等距的。如果因变量是二分的, 那么研究者就可以用logit、logistic、probit或者判别函数模型。logit模型有分类自变量。独立的顺序型变量和类别变量可以对容易解释的效应进行虚拟编码, 在logistic回归模型中同连续变量一起使用。在logistic模型里更喜欢用虚拟编码而不是效应编码来使解释变得简单。独立的连续变量可以简单地包含在logistic模型中。由于都用最大似然法来评估logit和logistic模型, 所以它们要求大样本。Probit分析要求满足正态分布假设, 然而probit系数比logistic回归系数更难解释。可以使用判别函数模型, 它假定误差协方差矩阵相同, 并且满足多变量正态分布的条件。可以用Box的M检验来评估误差协方差矩阵的等同性, 用Mardia检验来评估多变量正态性。只要它们最小特征值大于或等于1.0, 就可以得到区分因变量组的函数。如果因变量是顺序型的, 只有几个值, 应该使用具有累积logits的顺序型logistic模型。SAS、BMDP和LIMDEP程序都能实现累积logistic回归分析。

当因变量是一个多类别变量时, 可以选择多项式logit模型、判别函数分析或离散选择模型。在多项式logit模型里, 取其中一个类别作为默认类别, 参照这个默认类别形成logits。函数的个数比因变量的类别少一个。对判别函数的分析同样有这样的规律: 只有两种情况, 要么形成的方程比因变量的类别数量少一, 要么方程数量和方程所提取出的显著特征根一样多。在尽可能地区分组别时, 这些函数依据因变量属于哪一组而对个案进行分类。至于离散选择模型, 存在一套依赖类别的属性——如不同的运输方式——而这些属性在模型中是自变量。这些预测变量可能包括: 旅程的成本、运输方式的有效性和其他运输方式的成本。LIMDEP (logit) 和SAS (Genmod) 软件中有多项式logit分析的程序, LIMDEP还有一个离散选择模型的程序。

如果因变量代表一个可数的或者一个罕见的事件, 那么正确的分析是泊松回归。这个模型受到平均值等于方差这个特殊假设的限制 (Maddala, 1983)。SAS、SPSS和LIMDEP软件都可以进行这个分析。如果因变量的值底部、顶部或两端被截去, 那么LIMDEP (TOBIT) 或者SAS (LIFERG) 的tobit回归是合适的。

当因变量是连续的, 自变量也是连续的, 那么, 研究者可以进行多元线性回归分析。它的假设条件和ANOVA本质上是一样的。如果既有连续自变量又有因素作自变量, 那么, 这就是一个ANCOVA模型。除了回归和ANOVA假设之外, 还有一个前提是回归的同质性, 回归的斜率在不同水平的自变量上是相等的。这个假设可以由估计方差和自变量之间的交互作用的显著性来检验。如果这个相互作用是不显著的, 那么, 研究者就考察ANCOVA的最小平方或调整过的平均值。否则, 他或她不得不模拟单独的回归, 而且如果这些假设得不到满足, 那么就应该用非参数回归分析了。

多变量分析

当有两个或更多因变量时，就出现了多元变量的情形。Bock（1975）指出，有两种设计的方法可以产生多元变量：一是跟次数有关；二是跟测量有关。如果有人正在用多元方差分析（或叫MANOVA）的重复测量设计，多个因变量就是由多个试次构成的。通过组间平方和与cross-product矩阵的转换，可以调整重复测量之间的相关关系。所有的单变量一般线性模型的假设都适用于这些多元变量分析。但是，还需要检验另外一个假设。

如果多个因变量是连续的、相关的（但彼此不是重复观察），并且如果模型有一个或更多的独立因素，那么就用MANOVA。在这里，一个至关重要的假设是多变量的方差同质性，利用Box的*M*检验可以对这条假设进行评估。如果模型包含了作为多个因变量的重复测量，那么就要设定混合对称性假设；Mauchley的球形检验可以用来评价这条假设。如果这条假设不满足，研究者就可以用Grenhouse-Geiser或者Huynh-Feldt调整。其他单变量假设的运用，包括对残差的正态性，在多变量模型中也要进行评估。

如果多个自变量和因变量全都是连续的，那么就可以用多元回归分析。如果自变量是分类的，并且有一个连续的控制变量，那么可以用MANCOVA。除了MANOVA的假设，还必须满足回归的同质性假设（Tabachnick & Fidell, 1989; Tatsuoka, 1988）。

总 结

通过以上分析，研究者或分析者已能够对结果给出解释。这些统计结论的有效性依赖于选择适当的检验并理解它们的假设。这些假设也是需要检验的，它们的效度在面临相悖情境时，必须接受评估。如果假设不幸被违反了，研究者应该采取替代的分析方法，同时要鉴别它们的相对长处和短处；然后可以讨论诸如对当前主流理论的贡献、新知识的发展、空白的填补以及理论上矛盾的解决等问题。分析者因此而能理解和解释发现的结果，包括它们的优点和不足、提出尚未解决的问题和未来研究的方向。

推 荐 阅 读

在这里，我们为进一步学习的读者推荐几本书。对基础统计学，可以参考Kachigan（1986）或者Berenson、Levine、Rinddkopf（1988）。中等水平的读者，可以参考Snedecor（1980）或者Hays（1973）。如果准备进行深入研究，那么Cohen（1988）对效力和效应量的处理是必读的。关于研究设计和效度的经典著作包括Cook和Campbell

(1979)、Winer等(1991), 以及Keppel(1991)。其他有关相关性、多重回归和模型建立的复杂论述有Pedhazur(1982)、Cohen和Cohen(1983)、Draper和Smith(1981)以及Fox(1984)。有关一般线性模型和实验设计其他方面的优秀而复杂的论述请见Winer等(1991)和Kirk(1994)。这些书涵盖了ANOVA, 重复测量ANOVA、ANCOVA、MANOVA和MANCOVA。一般的多变量分析书籍包括Tatsuoka(1988)和Morrison(1976)。Kerlinger(1985)和Pedhazur & Schmelkin(1991)对测量、设计和分析作了全面论述。

至于更复杂的程序, 从事协方差结构建模的人应该阅读Bollen(1989)、Bollen & Long(1993)。学习有限因变量的同学可以参考Maddala(1983)和Greene(1993)。对非参数检验感兴趣的读者应该学习Seigel和Castellan(1988)、Daniel(1990)、Hollander和Wolfe(1973)或Conover(1980)。当样本总量不可避免地很小、样本总量严重失衡或者具有相等值时, 可以参考Mehta和Patel(1995)。学习了这些书, 读者就可以顺利地进行当代社会科学研究了。

参 考 文 献

- Anderson, A. B., Basilevsky, A., & Hum, D. P. J. (1983). Missing data: A review of the literature. In P. H. Rossi, J. D. Wright, & A. B. Anderson (Eds.), *Handbook of survey research* (pp. 414-494). Orlando, FL: Academic Press.
- Andrews, F. M., Klein, L., Davidson, T. N., O'Malley, P. M., & Rodgers, W. L. (1981). *A guide for selecting statistical techniques for analyzing social science data*. Ann Arbor: University of Michigan, Institute for Social Research, Survey Research Center.
- Bavry, J. L. (1991). *STATPOWER: Statistical design analysis system* (2nd ed.) [Computer program]. Chicago: Scientific Software, Inc.
- Berenson, M., Levine, D., & Rindskopf, D. (1988). *Applied statistics: A first course*. Englewood Cliffs, NJ: Prentice Hall.
- Bock, R. D. (1975). *Multivariate statistical models in behavioral research*. New York: McGraw-Hill.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley.
- Bollen, K. A., & Long, J. S. (Eds.). (1993). *Testing structural equation models*. Newbury Park, CA: Sage.
- Borenstein, M., & Cohen, J. (1988). *Statistical power analysis: A computer program* [Computer program]. Hillsdale, NJ: Lawrence Erlbaum.
- Borenstein, M., Cohen, J., Rothstein, H. R., Pollack, S., & Kane, J. (1992). A visual approach to statistical power analysis on the microcomputer [Computer program]. *Behavior Research Methods, Instrumentation, and Computers*, 24, 565-572.
- Cohen, J. (1978). Partialled products are interactions; partialled powers are curve components. *Psychological Bulletin*, 85, 858-866.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavior sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Conover, W. J. (1980). *Practical nonparametric statistics* (2nd ed.). New York: John Wiley.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin.
- Daniel, W. W. (1990). *Applied nonparametric statistics*. Boston: PWS Kent.
- Deming, W. E. (1960). *Sample design in business research*. New York: John Wiley.
- Dixon, W. J. (Ed.). (1990). *BMDP statistical software manual* (2 vols.). Berkeley: University of California Press.
- Draper, H., & Smith, N. (1981). *Applied regression analysis* (2nd ed.). New York: John Wiley.
- Fox, J. (1984). *Linear statistical models and related methods*. New York: John Wiley.
- Goldstein, R. (1989). Power and sample size via MS/PC-DOS computers. *American Statistician*, 43, 253-260.

- Greene, W. H. (1992). *LIMDEP user's manual and reference guide (Ver. 6.0)*. Bellport, NY: Econometric Software.
- Greene, W. H. (1993). *Econometric analysis*. New York: Macmillan.
- Hays, W. L. (1973). *Statistics for the social sciences*. New York: Holt, Rinehart & Winston.
- Hendry, D., & Richard, J.-F. (1982). On the formulation of empirical models in dynamic econometrics. *Journal of Economics*, 20, 3-33.
- Hettmansperger, T. (1984). *Statistical inference based on ranks*. New York: John Wiley.
- Hollander, M., & Wolfe, D. A. (1973). *Nonparametric statistical methods*. New York: John Wiley.
- Jöreskog, K., & Sörbom, D. (1988). *PRELIS: A program for multivariate data screening and data summarization*. Mooresville, IN: Scientific Software.
- Kachigan, S. K. (1986). *Statistical analysis: An interdisciplinary introduction to univariate and multivariate methods*. New York: Radius.
- Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Kerlinger, F. N. (1985). *Foundations of behavioral research* (3rd ed.). New York: Holt, Rinehart & Winston.
- Kirk, R. E. (1994). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Belmont, CA: Brooks/Cole.
- Kish, L. (1965). *Survey sampling*. New York: John Wiley.
- Leamer, E. E. (1983). Let's take the con out of econometrics. In C. W. J. Granger (Ed.), *Modeling economic series: Advanced texts in econometrics* (pp. 29-50). New York: Oxford University Press.
- Lenth, R. V. (1987). POWERPACK: A microcomputer package for power and sample size problems [computer program]. *American Statistician*, 41, 239-240.
- Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics*. New York: Oxford University Press.
- Mehta, C. R., & Patel, N. R. (1995). *SPSS Exact Tests 6.1 for Windows*. Chicago: SPSS.
- Morrison, D. F. (1976). *Multivariate statistical methods*. New York: McGraw-Hill.
- Nunnally, J., & Bernstein, I. (1993). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Pedhazur, E. (1982). *Multiple regression in behavioral research* (2nd ed.). New York: Holt, Rinehart & Winston.
- Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach* (student ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Rummel, R. J. (1970). *Applied factor analysis*. Evanston, IL: Northwestern University Press.
- SAS Institute, Inc. (1990). *SAS/STAT user's guide* (2 vols., Ver. 6). Cary, NC: Author.
- Siegel, S., & Castellan, N. J., Jr. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.
- Snedecor, G. W., & Cochran, W. G. (1980). *Statistical methods* (7th ed.). Ames: Iowa State University Press.
- SPSS Inc. (1989). *SPSS algorithms* (2nd ed.). Chicago: Author.
- SPSS Inc. (1990). *SPSS reference*. Chicago: Author.
- Tabachnick, B. G., & Fidell, L. S. (1989). *Using multivariate statistics* (2nd ed.). New York: Harper & Row.
- Tatsuoka, M. M. (1988). *Multivariate analysis: Techniques for educational and psychological research* (2nd ed.). New York: Macmillan.
- Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles of experimental design* (3rd ed.). New York: McGraw-Hill.



第 16 章

基本统计分析

DAVID N. DICKTER

MARY ROZNOWSKI

一般而言，我们不可能直接从研究的总体获得数据。所以，心理学家从样本中获得数据，根据这些数据推论总体。但是，这种结论有可能是偶然产生的，所以研究者有责任提供一些统计证据，表明研究中所报告的效应不是偶然的。因此，统计分析在心理学研究中是必不可少的。

基本统计分析是很广泛的内容，本章允许的篇幅有限，不可能对此内容进行深入彻底的讨论。因此，只能为大家提供一个概览，旨在提供一些分析程序方面的建议，鼓励读者作进一步探索。如要获取有关统计理论的更多信息，读者可以参考以下作者的书籍和文献，例如Hays（1991）、Lehman（1995）或Runyon和Haber（1991）。本章为读者提供了一个数据分析的良好开端，涉及心理学上使用的一些推断性统计，如 t 检验、方差分析（ANOVA）、相关和回归。此外，还讲述了针对频次数据的卡方检验。本章讨论的所有分析都是单变量的；多变量分析将在本书另外一章中讲述。

开 始

在做任何分析之前，研究者需要准备数据文件和挑选统计包。最常用的两种统计包是SAS和SPSS。它们使用相同的逻辑，并且都有十分友好的窗口界面。然而，研究者学习使用时需要参看用户手册。对于初学者而言，一方面，SPSS的手册比SAS更易理解。我们通常很难一下就弄明白SAS用户手册的内容，如果不懂程序的名称，就很难知道如何进行分析。例如，主要的回归程序在SAS中叫做PROC GLM，意为一般线性模型，而不是回归。另一方面，正如Levy和Steelman指出的（本书第17章），SAS提供了灵活的程序，供用户做数据处理，例如合并数据集等。如果读者想用SAS做基本数据分析，可参见基础统计分析《SAS系统手册》（SAS研究有限公司，1987）以获得更好的指导。另外，Kinnear和Gray（1994）著有关于基于Windows的SPSS的使用指南。

如果数据在纸上，那么研究者需要把它们输入电脑（此时更适合由助手来干这项工作）。一些统计包能辅助数据录入，但许多没有这项功能。一个数据文件只是简单的数据收集；研究者可以用文字处理或文本编辑器来生成这个文件。比较典型的情况是，每个被试只占数据文件中的一行或多行，数据记录从被试编号开始，接着是此被试在每个变量上的成绩。分类数据通常通过使用代码来节省空间。例如，头两行数据文件可能如下：

001 2 27 4 2 2 5 3 7 2 3 3 5 7 6 3 3 4 2 5 1

002 1 23 7 4 5 5 2 3 1 5 4 5 6 5 4 4 4 3 2 3

实验者应该保留代码表示哪些值的清单，称作代码表。在此例中，头三个数字是被试的号码。第二个字段代表被试的性别（1=女性、2=男性），第三个输入的字段是年龄，其他的代表被试对问题的反应。测量应使用Likert等级量表，从1~7级。

在建立数据文件时有一些事项需要注意。不要以列的形式输入被试的数据。因为统计包只认行的形式，转置行列十分困难，并且会带来不必要的麻烦。如有缺失的数据，在文件中用代码来代替，如SPSS中的“999”和SAS中的“.”（句点）。千万不要用空格来代替缺失值处，因为电脑会忽略掉空格（除非输入者另有指令），并且会将这一行中的下一个数据当做是缺失的那个。这样，电脑在这一行输入的数据将从此顺次弄错，就像一个学生在标准化考试答题时少填了一个圈，最终导致剩下的答案全部出错。如果一个人使用文字处理器来编辑文件，就能保证“只以文本形式”输入数据（也可称作“ASC II”或DOS文本）。否则，文件将包含无关字符，尽管我们在屏幕上看不到，但统计包还是会因此拒绝读取该数据文件。

一旦研究者创建了数据文件，那么在统计正式开始前，他（或她）必须使用统计包告诉计算机这些数字的意义（也就是说，给数值赋变量名）。数据准备在SAS中称为“数据步骤”或“输入步骤”，在SPSS中则称作“建立输入文件”。这一步骤包括用命令来指定数据点的位置。如果变量在输入时数据间没有空格（例如，如果被试001的数据写成“001227423537233576334251”），那么我们需要指定一个“固定的”格式，表明哪个变量归属哪一行。如使用所举例的正规数据格式，就不必做这一步。对于高级用户来说，SPSS和SAS允许使用许多更复杂的数据结构；用户手册中的指导有助用户写出合适的数据。

描述性统计

一旦写下读取数据文件的命令，研究者应该确认没有出现数据录入错误，并且更正那些不恰当的数值（例如，用程序或文字处理器来对变量值重新编码；根据Dollinger和Dilalla编写的本书第13章内容，对数据进行清理）。描述性统计具有这个功能。但重要的

是，描述性统计能够总结数据，并检验统计假设。

平均数和标准差提供了代表性分数数据变异量的重要信息。此外，研究者能用描述性统计和图表来检查数据集潜在的问题，包括非正态分布（一种远远偏离钟型曲线的分布）和离群值（大大异于其他大多数数值的分数）。使用数据图（如茎叶图或者正态分布图）是数据分析的好起点，用图形可以描绘样本数值的分布。前者列出了频次分布和离群值；后者在数据满足正态分布时是直线图形，当非正态时是曲线。偏斜度和峰度包含数据形状的进一步信息，由此可知数值是否聚合成偏离正态分布的特殊形式。偏态分布是不对称的，一边频次高，另一边频次低并带有长“尾巴”，峰度分布有高峰和两端的长尾。大多数统计包都提供简单的指令，用以产生有关数据集的信息（例如，SPSS中的描述性程序和SAS中的单变量参数检验程序）。

推断性统计

我们建议研究人员绘制数据图，并使用其他描述统计，以获得有关数据分布的更多信息。有关这一步骤的信息很少会单独出版，但是，如果学生想得到研究生助教奖学金，那么仅仅会描述性统计是远远不够的。研究者需要从样本推论总体。本章讲述了推断统计的几种类型。首先，我们引入关于统计显著性的讨论。

显著性检验

本章将以同样的形式讨论每种统计分析。首先，讨论每种检验的目标和潜在的假设，接着是对每种检验所用的统计作简短的评论，看看这些统计是如何决定应否推翻虚无假设。在讨论基本分析之前，我们想要留出部分空间来讨论Cohen（1990）和其他人提出的关于显著性检验的重要内容。

统计显著性检验首先假定虚无假设是真的（例如，实验操纵没有效应或两个样本总体平均数之间没有差别），并在此前提下推断，得到的结果为真的可能性。如果这个可能性很低（通常 p 值或者发生的概率小于5%），那么研究者应拒绝虚无假设，支持备择假设，说明实验操纵的效果确实存在。这个原理其实很简单，但常遭人们误解。Cohen（1994）指出，存在两种对假设检验和 p 值的误解。一种误解是， p 值等于0.05的“虚无假设是正确的概率”，其余的0.95，是研究假设为“真”的几率。其实，5%的概率指的是，如果虚无假设是真实的，就有5%的可能性得到观测到的数据；而这个概率不能直接证明一个虚无假设是否正确。Cohen（1994）概括出另一种误解，Rosenthal（1993）将其定义为“复制谬误”，是指有的研究者认为，如果在 $p < 0.05$ 水平上显著的研究结果上多次重复，研究者就会在大多数（95%）情况下得到显著性结果。不幸的是， p 值并不代表重复结果的几率。实际上，文献中效力水平（level of power）是一个更好地发现某个效

应可能性的估计（见Cohen, 1998; Levy和Steelman, 本书第17章）。

Cohen (1990, 1994) 和其他人也强调, 心理学研究应该超越占主流的显著/非显著二分法的评判标准。首先, 我们不应限制所有的研究都要以显著性检验为标准。具有理论价值的贡献不应被普遍的 $p < 0.05$ 的要求所遮蔽。严格坚持0.05可能会武断地否定那些未达显著性水平却有潜在价值的重要发现。在许多出版物中, Cohen和其他方法论学者认为, 研究者不应该只注意结果是否显著, 还要关注效应的大小。Cohen (1994) 对许多研究中出现的微小效应有微辞。其他研究者 (例如Schmidt, 1992; Hunter和Schmidt, 1991), 在他们的元分析研究中也注意到这个问题。元分析是一种可整合多个实验结果, 并计算出总效应的统计方法。研究者必须在心中谨记, 统计显著性和效应大小是不同的概念 (Rosenthal, 1993)。在 $p < 0.05$ 水平上, 显著的结果并不一定有“大”效应。同理, 研究者不应在小概率 (例如 $p < 0.001$) 出现时就声称效应“非常大” (Cohen, 1994)。

当我们报告一个显著性检验时, 应该报告其置信度, 它表明了虚无假设成立的范围。范围较大可能意味着有取样误差, 如样本容量不够大或者分数可变性低, 这些都提示我们, 即使虚无假设是错的, 我们也很难推翻它 (就是说, 可能隐含有严重的“二型” <Type II> 错误)。

总之, 除了假设检验之外, 研究者应作更多的事来得到有关数据的重要信息。此外, 只有意识到对假设检验的一般误解, 我们在下结论时才能避免常犯的错误。

统计检验的形式

统计检验遵循一定的步骤。统计指标可从研究结果中得出, 用以检验虚无假设。这种“检验统计”必须达到阈限值 (临界值), 这值最好在研究之前就定好。临界值是统计分布表中的一个数字。它取决于样本大小、研究者选定的 p 值和其他特性。 P 值由实验者对 I 类错误的容许度决定, I 类错误即研究者利用观察值错误推翻虚无假设的概率。如果统计量达到或超过了临界值, 那么研究者将推翻虚无假设, 得出有效应的结论。

t 检验

平均值差异的分析是一种基本的统计检验, 这检验有好几种不同的形式。我们在这部分将讨论两样本的 t 检验。假设一个研究者收集了实验组和控制组分数, 那么找出每组的平均测验分数就是一种比较组间差异的好方法, 即判断测验分数是否有显著差异。假设在一个40点量表上, 实验组20分, 控制组25分。这是偶然产生的细微差别, 还是效应存在的表现? 表述此问题的另一方式是, “两个样本来自同一总体还是两个不同的总体?” 通过 t 检验, 我们就能回答这个问题, 判断出两组分数间是否有显著性差异。答案不仅依赖于效应大小, 还依赖于样本容量和对 I 类错误的容许度 (α 水平)。

t 检验有一些基本前提, 研究者在做检验之前就应弄清这些假定是否成立。首先, 每

一组数据是否都符合正态分布？如果是小样本，检验则可能是不正确的，因为数据可能不符合统计检验的前提——即正态分布。然而，如果样本容量适中或较大（例如40~100，或更多），检验的分布越接近正态分布，检验将越正确（Hays, 1991）。其次， t 检验假定两组的方差基本相同。研究者需要先进行“方差齐性”检验。这个假设更重要，尤其是对样本数目不一致的小样本数据而言。第三，观察值间应互相独立，这样一个被试的分数就不会影响另一个被试的分数（例如，抄袭测验答案）。在做检验时，研究者应进行检查，确定数据是否符合统计前提。

t 检验用样本均数与方差之商作为统计量，将得到的 t 统计量与概率表中的临界值 p 值（通常是0.05，但也可能是0.01或0.001）相比较。如果 t 统计量达到或超过临界值，那么样本均数之间的差异就在选定的 α （I类错误的容忍度）水平上显著。需要注意的是，显著性检验分成“单侧”或“双侧”（也可分别称作“单尾”或“双尾”）。前者在事先假定特定组的平均数大于另一组时使用，而后者则在平均数不同、但并不作哪组均数更高时使用。对于双侧检验，为保证不增加随之而来的I类错误， α 水平应（还有 p 值）减半。结果，我们需要更大的 t 统计量来推翻虚无假设。

一种类似的 t 检验是成对样本的差异检验。例如，假设研究者想要比较两组职员的工作成绩差异，其中一组接受了为期一周的职业训练课程，另一组则没有接受。随后，记录每一个职员在训练前和训练数周后的成绩。在此测验中，我们假定每一对观察值（训练前/训练后）的配对独立于其他对的。为了检验训练组的成绩是否比控制组高，研究者可对两组被试训练前后的差异进行单尾 t 检验。

方差分析（ANOVA）

t 检验适合检验两组均数是否存在显著差异。然而，检验中比较组的数量越多，I类错误的几率越大。例如，如果研究者做5个 t 检验，并允许每个有0.05的I类错误，那么错误率将达到 $1-0.95^5$ ，或有接近23%的错误几率拒绝虚无假设。多次 t 检验的替代方法是方差分析。像 t 检验一样，方差分析是对各组均数的比较。而与 t 检验不同的是，它不限制比较组的数量（当然，组间的比较应该基于理论预测，而不是随机地在从不同的组之间进行）。研究者用方差分析检查组间和组内分数的变异性，同组被试分数的变异由个体差异和随机误差造成。但如果有效应的话，组间差异就会更大。当只有两组做比较时，用方差分析和用 t 检验会产生同样的结果。

与 t 检验一样，方差分析假设各观察值之间是相互独立的，数据呈正态分布，组间方差齐性。方差是否同质会影响到结果，尤其是当样本大小不等的时候。因此，应该格外注意这种影响。

方差分析的第一步是，决定是否有一组的平均数与其他各组显著不同。这就是 F 检验。如果没有差异（也就是说如果 F 检验不显著），那么就没必要再去比较任何两组之

间的差异了。这个结果提示研究者应该接受虚无假设，承认各组平均数相等。另外，如果整体 F 检验显著，研究者应该做“事后”组间平均数的比较。一个显著的 F 值表示至少有一组的平均数显著异于其他各组。一般，研究者应该检验他（或她）对组间数据的假设。例如，研究者可能假设A组的平均数显著异于B、C、D组，或A和B组的平均数显著异于C、D组的平均数。重要的是，这些统计检验必须基于研究假设（例如，A组和B组将比C组和D组具有更高的测验分数，因为A组和B组接受了一种与C组和D组不同的指示）。很清楚，我们不应该在观察过所有平均数之后找到一个显著异于它组的均数，然后对它进行检验，“找出”一个显著的结果。研究者必须在看到数据之前就决定做组间比较，而不是在看到结果后才做假设，声称结果正是他们所预期的。

在发现显著的 F 值之后，研究者能用多种不同的统计检验来比较组间差异。每种检验都会使研究者承担起犯I类错误风险。Tukey和Scheffe检验风险最小，Fisher和Newman-Keuls检验风险较大些。然而，如果用Tukey和Scheffe检验，I类错误的减少会换来II类错误增加的代价，即错误推翻实际存在的效应。尽管如此，一般来说，Tukey检验还是最好的选择。（Keppel, 1991）。

方差分析还有更复杂的内容。读者应查询其他的资料（例如Keppel, 1991），以获得更多的信息。因素方差分析，包括两个以上的独立变量，可以计算出交互作用。一组变量的结果在另一组变量不同水平下存在差异，就产生了交互作用。例如，如果在实验中，学生为自己设定了高目标，那么他们能比给自己设定低目标的学生解决更多的物理问题，但只有在使用高级认知能力时，我们才能观察到这种差异，这就是说自我目标设定与能力发生了交互作用。交互作用提示，在没有确定学生是否具有高级认知能力之前，我们不能光凭自我目标水平的高低来归纳对物理问题的解答。自我目标水平和能力的“主”效应取决于学生其他的变量。

本节的篇幅简短，希望读者能从更多的渠道获得有关方差分析的知识（对本章描述的其他分析方法也是如此）。特别是方差分析允许重复测量设计，即被试接受一种以上的处理，研究者则测量他们之间的差异。Keppel（1991）的手册给出了有关使用重复测量和其他实验设计进行方差分析的信息。

相关

在讨论回归之前，有必要花些精力了解相关。相关是指两个变量的联系程度，在+1.0和-1.0之间取值。如果两个变量正相关，那么一个变量增加，另一个也增加。如果两者是负相关，那么一个增加，另一个则减少。如果它们没有联系，相关即为零。如果每个被试在两个变量上的得分绘制成xy轴间的点（也就是说，用“散点”来表示），零相关会大概表现为粗略的圆形区域。正相关则接近线性递增形式，负相关接近线性递减。相关可用于推论总体两个变量的关系，例如检验两个变量无关这样的虚无假设。在做出推

论前，数据需假定符合正态分布。

尽管习惯上的虚无假设是总体相关为0.0，在实践中样本之间的相关为0.0的情况并不常见。然而，就算是微小的相关也可能是显著的，因为相关的显著性部分依赖于样本量。的确，给定一个大样本，理论上不相关的变量在0.001的相关系数水平下也会达到显著水平，这种现象称为“杂质因素”（Meehl, 1990，由Cohen, 1994引用）。心理学家一般只看重 ± 0.3 以上的相关或更高。通过将变量A和B间的相关系数进行平方，可知A的变异在多大程度上可被B值所解释（反之亦然）。因此A和B之间0.3的相关意味着A可以解释B 9%的变异。

千万不要仅靠相关系数而对两个变量间的关系轻易下结论。人们通常会犯这个错误，如声明美国SAT分数的下降，是因为美国青年花更多的时间看电视。相关并不意味着两者之间有因果关系。相关所能说明的只是有正或负的某种联系，并不能推出A导致B或B导致A。

回归

前面讨论的“能被解释的方差”这一概念对回归分析来说很重要，这里简短讨论一下。研究者可以用回归分析来做方差分析能做的所有事。方差分析考察的是离散变量的组间差异，但如果变量不是离散的而是连续的，例如年龄，那怎么办？研究者可以将样本分成两个离散的组，例如，40岁以上和40岁以下，比较这两组的差异。但这样做会将5岁和37岁的放在同一组中。最好的方法是不用2、3、10或其他分类数字，而利用年龄变量本身的连续性。回归能做到这一点；方差分析则不行。经典的回归包括制造一个线性方程，预测独立变量的分数。方程代表一条最佳的穿过散点图用以描述自变量和因变量关系的直线。方程中的 β 权重或自变量的系数是自变量和因变量关系的有关信息。例如，我们可以这样理解回归方程因变量的 β 权重：想像xy坐标中，绘制的点是被试分数在自变量和因变量下被试的分数。如果一个人画一条简单的直线来最佳拟合这些数据，那么直线的斜率（就是 β 权重）代表了自变量每个单位上的变化所造成的因变量上的变化。

正如 t 检验和方差分析，回归分析的假设是各变量间独立、正态分布、变异恒定。此外，它还假定自变量和因变量之间有典型的线性关系。但研究者可能会想通过测量数据间的非线性趋势，来获得更好的回归方程，表达这些数据（见Cohen和Cohen, 1983）。

简单的线性回归是用单个自变量来设计因变量的分数。回归可确定一个连续自变量（例如，GRE分数）所解释的因变量的变异（例如，研究生第一年的学分成绩点）是否有显著性。为此，可计算两者之间相关系数的平方（ R^2 ），并检验它是否显著大于零。如果确实如此，那我们就可以得到如下结论：“GRE成绩能解释第一年研究生成绩方差的16%（如果 R^2 是0.16）。”这样，回归分析给出了一些描述自变量和因变量联系程度的指标。另一种检验自变量显著性的方式是用 t 检验来检查 β 权重的估计值（也称为“参数

检验”），并决定结果是否显著不等于0。

多重线性回归可以用来确定：一组自变量解释的因变量中的变异是否显著不等于0。通过因素方差分析，我们可以发现交互作用，因为自变量数量已不只一个。此外，可能有人会提出一些更复杂的问题，例如一个自变量在多大程度上可以独立解释自变量变异。也就是说，一个自变量比其他变量来说，能在多大程度增加解释的力度？这种分析方法叫做方差半相关。注意，将变量以不同顺序引入回归方程会产生不同结果。读者应参考其他文献资料，如Cohen和Cohen（1983）获得有关排除变异和使用回归步骤（例如前向选择、后向消除和逐步回归，来评价变量贡献的信息）的信息，这些手段能帮助我们评价一个特定变量或一组变量的可能贡献。

卡方检验

另一种研究涉及分析频次，来看看变量是否在某范畴上存在差异。例如，一个人可能想研究男性和女性是否对社会问题持有不同的观点（如在餐馆是否应禁止吸烟）。上面讨论的统计方法，没有一种适合研究这类问题，因为这些数据是分类的。此外，这种研究检验的是各组人（例如，回答“一直”、“有时”、“从不”的人群）的整个频次分布，而不是一些概要的统计信息，例如平均数和方差。在这个例子中，卡方分析能回答“男性和女性在这个社会问题上是否持不同意见”。如果支持禁烟的男性数和支持吸烟的女性数之间有显著差异，那么，虚无假设（两者观点之间没有差别）将被拒绝。

与其他统计方法一样，进行卡方分析时要做一些假设。一个假设是各项观察之间是独立的。也就是说，一个被试对吸烟禁令问题的反应不应被另一被试的反应所影响。此外，观察值应该落在相互排斥、相互独立的类别范围中（例如，吸烟者和不吸烟者）。最后，一般来说，某个特定类别的预期频数不应该小于5（见Hayes, 1991）。

卡方分析运用观察的和预期的频数来检验虚无假设。在虚无假设之下，应有相同比例的男性和女性支持（或反对）吸烟禁令。男性和女性支持（或反对）的预期频次应该满足这个假设。观察（真实）频次与预期频次之间相似或者相异的程度就将成为接受或者推翻虚无假设的证据。

卡方检验遵循与其他统计显著性检验相同的形式：如果卡方值超过了阈限，那么研究者会拒绝组间无差异的虚无假设。这就意味着男性和女性对吸烟禁令持不同意见。卡方分析对调查研究尤其有利，因为这些调查使用频次数据，本章介绍的基本分析方法对此无能为力。

总 结

通过以上对卡方检验、*t*检验、方差分析以及相关和回归所作的概述，我们希望能回

答读者的一些问题，并帮助读者去发现更多的问题。当然，本章这个类似菜谱的东西，不能作为大厨或个人正式训练材料的替代品。但是，可以试试本章所述的方法，查阅引用的文献，与同事、教授和导师交流协商。掌握了这些基本的统计分析方法，就等于掌握了无价之宝。你将在研究中经常用到它。

推 荐 阅 读

鼓励读者根据本章讨论的每一个主要统计议题寻找更多的资料（例如，Cohen和Cohen，1983，相关/回归；Hays，1991，卡方检验；Keppel，1991，方差分析；Runyon和Haber，1991， t 检验）。关于统计和SPSS（例如，Bryman和Cramer，1994；Rose和Sullivan，1993；Stevens，1990）或SAS（例如Stevens，1990）使用新文献也有帮助。

很自然地，需要再次提到统计软件包公司（例如，SAS研究公司，1993；SPSS公司，1993）出版的“基础”手册。SAS和SPSS都为它们的主机和个人电脑配置了用户手册，通过大学书店可以买到。另一些手册能作为辅助材料使用，例如Spector（1993）的SAS程序语言和数据处理以及Kinnear和Gray（1994）的SPSS指南。

参 考 文 献

- Bryman, A., & Cramer, D. (1994). *Quantitative data analysis for social scientists* (Rev. ed.). London: Routledge.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, 45, 1304-1312.
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 49, 997-1003.
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavior sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Hays, W. L. (1991). *Statistics* (5th ed.). New York: Holt, Rinehart & Winston.
- Hunter, J. E., & Schmidt, F. L. (1991). Meta-analysis. In R. K. Hambleton & J. N. Zaal (Eds.), *Advances in educational and psychological testing: Theory and applications* (pp. 157-183). Boston: Kluwer.
- Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Kinnear, P. R., & Gray, C. D. (1994). *SPSS for Windows made simple*. Hillsdale, NJ: Lawrence Erlbaum.
- Lehman, R. S. (1995). *Statistics in the behavioral sciences: A conceptual introduction*. Pacific Grove, CA: Brooks/Cole.
- Rose, D., & Sullivan, O. (1993). *Introducing data analysis for social scientists*. Buckingham, UK: Open University Press.
- Rosenthal, R. (1993). Cumulating evidence. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 519-559). Hillsdale, NJ: Lawrence Erlbaum.
- Runyon, R. P., & Haber, A. (1991). *Fundamentals of behavioral statistics* (7th ed.). New York: McGraw-Hill.
- SAS Institute, Inc. (1987). *SAS system for elementary statistical analysis*. Cary, NC: Author.
- SAS Institute, Inc. (1993). *SAS user's guide: Basics*. Cary, NC: Author.
- Schmidt, F. L. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist*, 47, 1173-1181.
- Spector, P. E. (1993). *SAS programming for researchers and social scientists*. Newbury Park, CA: Sage.
- SPSS, Inc. (1993). *SPSS for Windows: Base system user's guide, release 6.0*. Chicago: Author.
- Stevens, J. P. (1990). *Intermediate statistics: A modern approach*. Hillsdale, NJ: Lawrence Erlbaum.

第 17 章

高级统计方法的使用

Paule E. Levy

Lisa A. Steelman

一个特定的研究中应该采用何种统计方法，关键在于要研究的问题。Tabachnick 和 Fidell (1996) 认为，研究问题可以分成四类。我们将采用他们的分析框架，以求更好地理解和使用高级统计。本章会涉及这四类，但重点会放在其中三类：变量之间的相关程度、组间平均数差异以及结构。在某一个特定的研究中，研究者可能会问及其中一类或所有类别的问题，这样，对于某一个特定研究，就会有来自不同统计分析方法的汇报结果。在进入高级统计技术之前，我们先谈谈统计效力，因为这是一个在研究设计中最重要也最常被忽视的问题。

效 力

下面我们将讨论影响统计效力的因素，读者可参阅Cohen (1998) 对效力的完整的、技巧性的讨论，或者参阅Lipsey (1990) 对这个问题更具可读性的论述。统计检验的效力是“当虚无假设错误时，拒绝虚无假设的可能性”。也就是说，统计效力是指当两个实验组之间存在差异时，正确地拒绝虚无假设（认为两组没有差异）的概率。

四个因素将决定统计检验的效力：(a) 统计检验自身；(b) α 水平；(c) 样本大小；(d) 效应量大小。首先，在任何数据分析程序中，我们都希望能选择对待检验假设最为敏感的统计方法。对于同样的数据，不同的统计检验有不同的效力。例如，当满足参数检验的假设时，非参数检验（例如，曼惠尼 U，Kruskal-Wallis）一般比参数检验（ t 检验，ANOVA）效力要小。违背统计检验假设会降低检验的效力，也限制了研究发现的可靠性。另外，检验非方向性的假设（双尾检验）比检验方向性的假设（单尾）效力小。因此，研究者应谨慎选择统计检验方法。

决定统计效力的第二个指标是 α 水平。它表示我们错误地拒绝虚无假设（即虚无假设其实是成立的，而我们拒绝了它）或者犯 I 类错误的概率。而 β 水平表示我们错误地

接受了虚无假设（即虚无假设其实是不成立的，而我们接受了它）或者犯Ⅱ类错误的概率。这两个水平隐含着难以权衡的冲突：犯Ⅰ类错误和犯Ⅱ类错误的比率存在相关。例如，一个研究者选择了一个非常小的 α 水平，如0.001，这样他很难犯Ⅰ类错误（即他或她不太可能拒绝虚无假设），但 α 过小会使 β 膨胀，并增加犯Ⅱ类错误的可能性（即他很可能接受错误的虚无假设）。研究者应该权衡自己的研究对两类错误的可接受度，从而设定 α 。在行为研究中，我们通常关心的是Ⅰ类错误，因此会把 α 水平设置得很低（0.01或者0.05）。

效力被定义为 $1-\beta$ ，它与Ⅱ类错误直接相关。这当然也意味一个检验的效力间接和 α 联系。正如我们上面提到的： α 和 β 相关。换句话说，上面我们提到的权衡会直接影响检验效力。因为效力是错误地接受虚无假设的可能性，所以任何Ⅱ类错误的增加（接受错误的虚无假设）都会降低效力，而增加Ⅰ类错误会增加效力。这说明小的 α 水平会降低效力。提高效力的一种途径是提高允许犯Ⅰ类错误的底线（设置大一点的 α 值）。前提是研究者必须针对两种错误类型对其研究的重要性做出一个恰当的判断，然后设定相应的 α 水平。

样本大小也是决定统计检验效力的一个关键因素。样本越大，统计检验越可能探测到真实的效应。因为当样本比较大时，如果取样恰当，就能比较好地代表总体。但人们普遍存在一个共同的误解，就是只要增加样本的大小都会增加效力。Schmidt、Hunter和Urry（1976）证明，在确认性研究中，要想为满足效力而需的样本大小，一般比大家认为的要大得多。因此，如果你希望通过增加样本大小来提高效力，就需要增加足够多的样本量，然而，这种提高效力的方法其代价非常大，而且非常耗时。

最后，来看看效应量。效应量的大小表明了实验组和控制组之间的差异。而且，效应量的大小取决于两组被试的均值和组内变异。两组被试之间的效应量越大，效力就越大。效应量和实验设计相关，因为实验设计的错误越多（如抽样错误、测量错误、操作自身的质量不高）、效应量越小，导致效力越低。因此聪明的实验者在设计开始时，就应该尽量减少外来错误，通过影响效应量来提高效力。我们应该清楚，效力非常复杂，而且不能充分概念化，或仅仅通过关注研究中的单一成分来控制，而必须将实验设计视为一个整体来对待。

由于效力 and α 水平、样本大小、效应量相互关联，所以一个研究的效力在很多方面都是有用的。在效应量给定的情况下，效力分析可以用来决定样本大小，也可以进行事后效应分析，来确定研究的效力大小，或者是探测一个既定样本的最小的效应量。Lipsey（1990）对效力进行了更细致的分析，并提供了一些指导和建议，读者可以用来参考。现在我们使用Tabachnick和Fidell（1996）的方案来讨论各种不同的统计方法。

变量之间的相关度

Tabachnic和Fidell (1996)的第一部分讨论了变量之间关系的研究问题,并提到使用相关和回归方法。本书第16章,讲述了二元或多元变量的相关、回归和卡方一致性分析。典型相关(canonical correlation)是线性回归方法的一个扩展,它考察多个预测变量/自变量和多个标准变量/因变量之间的关系。例如,研究者可能对社会经济状态指标(如收入、工作水平、职业地位、财产)与生活质量感(如工作满意度、家庭满意度)之间的关系感兴趣,前者为多个标准,后者为多个待预测的变量。典型相关方法可以告诉研究者两组变量之间是否有统计上显著的关系以及是怎样的关系。我们知道,在二元或者多重相关中有一个表达特定相关关系的指标。在典型相关扩展中,可能存在多对典型变量(canonical variates),表示其更为复杂的关系。Boko (1990)在他的多元变量相关分析一章中,详细阐述了典型相关,可读性很强。他指出,实际上大多数基础的、熟悉的统计方法,如简单和多元相关、多元回归等都可以被看做是典型相关的特例。其他的信息请参见Stevens (1992)、Tabachnic和Fidell's的著作,他们讲述了这种方法的概念解释和技术细节。

组间差异的显著性

大多数的实验研究感兴趣的是不同条件下的不同被试在某个因变量(dependent variable, 缩写为DV)上的差异,同时,组间差异也是大量对自然状态下形成的团体进行实地研究的核心。我们应该清楚,组间差异的显著性是最常遇到的研究问题。Dickter和Roznowski (16章)讨论了平均差异的基本检验,如 t 检验单因素方差分析(one-way ANOVA)、协方差分析(ANCOVA)、因素性方差分析(MANOVA)和因素性协方差分析(MANCOVA)。本章主要关注各种分析方法的多元扩展。我们会讲述如何运用这些扩展的方法,对多变量的组间差异进行研究。

Hotelling's T^2

一个发展心理学家可能会对10岁和12岁儿童在多个认知功能指标(如IQ测验、阅读测验、一系列的皮亚杰任务、实践知识)上的差异感兴趣。这些数据可以通过使用Hotelling's T^2 分析。这是 t 检验的多变量扩展。Hotelling's T^2 告诉研究者两组被试是否在综合因变量(DV)上有差别。这些数据可以使用4个分离的 t 检验分析,但考虑到实验误差概率累积的问题,我们强烈反对用这个方法。误差概率累积使Ⅰ类错误膨胀。换句话说,即使个体差异(分离的 t 检验)使用 $\alpha=0.05$ 的标准检验,多重检验后,实验误差就可能是

0.20（关于这个问题更详细的讨论见Kappel, 1991）。研究者还认为 α 水平是0.05，但实际上它已膨胀到0.20，Hotelling's T^2 通过一次综合因变量的 t 检验避免了这个问题。

One-way MANOVA

一个工业/组织心理学家可能对不同的领导风格感兴趣（改革型、事务型、独裁型）如何影响多个行为变量（缺席率、协作程度、产量、产品质量）。因为在这个例子中有3组而非两组，Hotelling's T^2 检验是不适合的。在单变量的分析中，这些数据可以使用ANOVA（而非 t 检验）分析。例如，如果我们仅对领导风格、产品质量的影响感兴趣，上一章（第16章）里的知识会告诉研究者如何使用one-way ANOVA。但对多重变量的情况，应该使用MANOVA（多元方差分析、因素性方差分析），因为此处我们有多重DV（领导风格如何影响多个行为变量），所以这里需要使用one-way MANOVA。

MANOVA考察两个以上总体在两个以上综合的DV上的差异。在以前的例子中，我们指出做多重 t 检验会导致I类错误的膨胀；如果我们选择对每个DV做ANOVA，也会产生同样的问题。机敏的研究者会采用更适宜的分析方法—MANOVA来避免I类错误的膨胀。这种方法还有其他好处。首先，可以扩大研究项目的范围（Spector, 1977）；其次，使用多重测量使得研究者可以检查不同DV之间以及自变量与DV之间的关系。

大量统计文献和杂志文章详尽介绍了MANOVA的操作技术（Bray & Maxwell, 1982, 1985, Spector, 1977; Steven, 1992, Tabachnic & Fidell, 1996）。尽管我们的讨论特别注意借鉴这四个资源，但我们将尽可能地避免统计和计算的问题，而是集中讨论概念。

MANOVA的操作分两步。第一步经常称为总体检验，它集中考察各组之间在综合的DV上是否存在差异，这是ANOVA中的总体 F 检验的推广。在我们的例子中，它回答的是领导风格是否影响行为变量。通常，对这个问题的确定性的回答（我们拒绝虚无假设，认为不同的领导风格下的行为变量是没有差异的），仅仅是研究的一部分。显然，我们还想知道差异在哪儿。这个简单的问题——差异在何处，通常是统计方法的焦点，也是研究者真正的兴趣所在。因此紧接着的第二步是检查这些差异。

在得到显著的MANOVA结果后，研究者将从多方考察差异所在。建议读者参考上面列出的资源，更详尽地了解不同方法的利弊和不同方法的操作过程。我们在这里花一些时间讨论两个最普遍的方法。

最常用的方法是单因素的ANOVA，在上例中，我们会执行四个ANOVA（对4个DV）。假如MANOVA考察的多重效应仅仅体现在缺席率和作业成绩上，ANOVA在这两个DV上将显著，而在其他两个DV上不显著。当然，ANOVA可以告诉我们组间差异何在。但是，它没有告诉我们哪些组在这些变量上差异显著。如果我们只有两组，这不成问题，因为差异只能存在这两组之间，但是我们有三组，研究者必须进一步使用事后检

验 (post hoc test) 探讨 (参阅 kappel, 1991, 有助于对这个检验更细致的理解), 如哪两组 (1 vs.2、1 vs.3、2 vs.3) 在缺席率上有差异。Specter (1977) 认为, 使用ANOVA后续程序的的最主要益处在于假设检验。我们曾强调要研究的问题决定我们的统计方法。Specter (1977) 建议研究者应根据研究的问题来选择MANOVA进行显著性检验。如果问题只需要检验在哪个DV上有差异, 单变量ANOVA是最好的方法。

第二个常用的方法是使用鉴别分析 (DA, discriminant analysis), 它可以找出哪些DV的线性组合可使组间的差别最大 (Specter, 1977)。为有助理解, 可以从数学上将把鉴别分析当成一种回归, 组成员类比为标准变量。这样, 鉴别分析告诉研究者, 如何根据数据预测成员所属的组。在关于领导风格的例子中, 我们可知如何用效绩来预测个体处于何种风格的领导之下。判别系数和回归权重相似, 揭示了各个DV对最大区分函数的贡献。Specter (1977) 认为, 因为研究者是对预测和分类感兴趣, 提示使用DA作为随后分析就比较恰当。假设我们的MANOVA是显著的, 我们就可以得出结论说: 领导风格影响效绩。如果我们想要知道哪个DV或者哪些DV的线形组合可以把组成员区别得更好, 鉴别分析应该是回答这个问题的最好方法。

Tabachnic & Fidell (1996) 把鉴别分析看作是针对他的第三类研究问题的主要分析方法中最主要的一种, 它可以用来预测组的成员关系。尽管在这一章, 我们把DA看做MANOVA的第二步分析, 我们仍要强调, DA其实是独立的分析工具, 感兴趣的读者可以参看Tabachnic和Fidell的著作 (1996)。

一个MANOVA的例子

当然, 对有些研究项目, 使用单变量ANOVA和鉴别分析作为MANOVA的事后分析是很好的做法。我们将举一个应用心理学文献中的例子来解释这个方法。Abelson (1987) 对能够预测雇员是否会辞职的变量感兴趣。他认为, 个体离开的原因可以分成两种: 第一种是组织不能控制的, 如伴侣在另一个地方得到了一份工作, 这些是不可避免的辞职者; 第二种是组织可以控制的, 如雇员得到另一份有更好条件和薪水的工作, 这些是可以避免的离开者。最后, 有的个体不愿意离开, 称为停留者。Abelson的文章称, 停留者和不可避免的离开者类似, 同时这两个群体和可以避免的离开者不同。理由是无论留下还是不可避免地离开, 都是不想离开, 就算离开也不是由于组织的原因, 而可以避免的离开者却是因为不同的理由而离开组织的。

Abelson (1987) 使用看护人员对一个组织的情况进行了研究, 感兴趣的变量包括雇员的年龄、婚姻状况、工作满意度、组织认同、工作压力、对领导敏感性的知觉。采用MANOVA进行显著性检验。和预期一样, 组间差异在综合的DV上显著。接下来使用单变量的ANOVA检验差异所在, 换句话说, 组之间在哪些变量上有差异。结果表明, 在接近一半的变量上, 差异是显著的, 包括工作满意度和组织认同。这揭示在这些变量上组

间存有差异,但我们不知道哪些组在这些变量上有差异。Abelson预测不可避免的离开者和停留者相似,这两组与可避免的离开者那一组间会存在差异。和预期一致,不可避免的离开者和停留者,在工作满意度和组织认同(以及其他的变量)上没有差异,但是两者和可避免的离开者间却有差异,后者满意度相对低,组织认同感低。

除了使用单变量的ANOVA方法外,Abelson(1987)也使用了鉴别分析,试图判定哪些变量可以预测停留者、不可避免的离开者和可避免的离开者(p.383)。鉴别分析表明,6个变量能够达到最好的区分效果,可以把可避免的离开者与停留者、不可避免的离开者都区分开。鉴别分析的击中率达80%,这意味着判别函数(也就是说,个体在重要变量上的分数)把80%的被试归类到正确的组中去。我们可以看到,MANOVA的结果、单变量ANOVA和事后检验、鉴别分析都支持了Abelson事前的预测。

其他的扩展

我们要指出,这里只是稍微开启了通向多重DV的分析世界大门。与ANOVA一样,也可以把MANOVA扩展到只有多个自变量的因素设计。同样,MANOVA是ANOVA的多变量扩展,可以在控制可能与因变量相关的其他变量基础上,考察一个自变量对多个DV产生的效应。例如,在我们上面提到的领导者的例子中,一个领导人在某个岗位上的工作年限可以成为一个单向ANOVA中的协变量,这样就变成了一个单向MANCOVA。请记住,协变量设计的重要目的是通过控制协变量与DV间的关系来降低误差(见Pedhazur & Schmekin, 1991, 关于正确使用和误用协变量分析的讨论)。

结 构

Tabachnic 和 Fidell (1996) 讨论的最后一类关心的问题是一组变量下潜在的结构,常用的分析工具包括主成分分析和因素分析。这些方法经常用来减少变量个数,或者判断多重变量的潜在结构。例如,主成分分析可以用来把被试提出的100个项目(针对他们如何不喜欢他们的工作)减少到一个可用的维度数目,如减少到不喜欢他们的上司、工作环境、与同事有关的、工作内容、薪水这5个维度。这5个独立变量可以用在随后的分析中,而非在100个原始变量上使用不同的分析。因素分析可用来检验研究者测量一个结构的10点量表实际上测量的是一个还是多个结构。换句话说,因素分析可以用来证明量表的有效性以及这个特定结构是单一维度还是多维度的。

路径分析是多元回归的扩展,它也可用来处理数据的结构。当研究者分析变量间特定的因果关系时,路径分析是有用的。例如,Levy (1993) 使用路径分析考察一个预测自尊、心理控制源、自我评价和归因之间关系的模型。这个分析是带有少量额外考虑的多元回归的直接扩展。读者可以参考Pedhazur (1982, 第15章)、James、Mulaik和Brett

(1982)以获得更详细的讨论。一般来讲,每个DV在对DV有影响的每个独立变量上都有回归。例如,Levy(1993)预测,自我评价会受到自尊和心理控制源的影响,归因受自我评价和心理控制源的影响。对每个DV进行多元回归。回归中的权重表示变量间关系的方向和强度。可以通过重现计算变量间的相关以及通过Q检验来检验模型与数据的拟合度,Q是传统的拟合度指标(见Pedhazur,1982)。如果模型看起来与数据拟合得较好,研究者即可得出结论说数据和前面的因果模型一致。请注意:这不等于说路径分析可以证明因果关系(see Bobko,1990)。路径分析不可能做到这点。

社会科学中发展最快的分析方法之一是结构方程(SEM, structural equation modeling)。Pedhazur(1982)在他对SEM的介绍中,把它视为分析因果关系的方法的总称(注意,路径分析是其方法之一)。他也指出,比路径分析更有效的方法,例如LISREL建立在更少的限制假设之上。SEM的方法之一LISREL代表着“线性结构关系”,是由Jöreskog在1970s提出(参看Jöreskog-Sörbom,1993最近的统计程序版本和相关的讨论),它建立在最大似然统计理论基础之上,而不是典型的最小平方统计理论基础之上,最小平方统计理论是简单和多元回归的基础(这已在前面一章讨论过)。显然,结构方程(LISREL是其方法之一)非常复杂,我们这里不能给出解释。除了Pedhazur(1982)、Joreskog和Sorbom(1993)的介绍外,也可见Bobko、Long、Pedhazur和Schmelkin(1991,第24章)。

Bobko(1990)对LISREL做了一个很好的总结,他指出,LISREL是路径分析(结构方面)和因素分析(测量方面)的一般性扩展和整合。实际上,在一篇高引用率的文章里,Anderson和Gerbing(1988)提出LISREL分析分为两个步骤:第一步与因素分析有关,通常叫做考察测量模型;当确定潜在结构后,第二步与结构方面有关,与路径分析类似,用来检测结构,被称为检验结构模型。LISREL模型处理的是潜变量(即假设结构)和显变量(即假设结构的外示)。例如,一个教育心理学家可能对下面的潜变量之间的关系感兴趣:社会经济背景、学术成就、工作成就。假设社会经济背景可以通过家庭收入、地区内小学每个学生的开销、父母的职业地位来测量,学业成就可以通过接受学校教育的最后程度的GPA、荣誉/奖金和在学校中的排名来确定,而事业成功就可以通过收入、职业地位、组织内的工作地位和同伴对个人成功的评价来确定。所有这些测量指标都是这个教育心理学家感兴趣的的结果的表现。这些测量模型检测的是测量能在多大程度上确实代表了假设的结构就是对测量模型的检验。分析的结构部分检验的是潜变量(通过显变量测量)是否以研究者预期的方式发生关系。例如,我们可能预测社会经济背景与学业成就有因果联系,且这两个变量都对事业成功有直接影响。

综上所述,LISREL估计每个标志变量在各个潜在变量上的负载,以此提供有关测量模型的信息。换句话说,这些是否为每个结构的测量?如果不是,在检验结构模型之前,需要对测量模型做更多的工作;如果数据支持测量模型,研究者可进一步考察结构

模型，以LISREL考察潜在变量之间的因果关系。我们应该使用一个或多个拟合度指标，来检验测量模型和结构模型。注意，这里只是粗略地介绍了一个复杂而有用的方法——LISREL。

总 结

我们愿再提供一点关于统计计算机程序的信息来结束本章。尽管我们的统计分析已经变得更为复杂，学生仍可以享受知识的乐趣，因为我们的统计计算机程序已经变得更为友好。手工进行因素分析的时代已经结束，被大量的建立在PC机基础上的软件分析程序所代替。我们在最后一节简单地介绍以下最为通用的计算机程序。

首先，有三个大家都知道的计算机应用软件：SPSS、SAS和BMDP。SAS和SPSS在社会学和行为科学中得到了广泛的应用。我们的经验是，SAS在处理大量的数据组（如整合数据组或者转移数据）时表现更好。但在这章和前一章讨论的统计分析实际操作时，SPSS和SAS都很出色。每个程序都十分全面，你很容易就会发现这两个软件（也许还有BMDP）。你可以咨询大学里的计算机技术人员，以获得更多应用软件的信息。

数据分析的新方法是基于PC环境的。一些程序，如SYSTAT和MYSTAT是专门为PC环境所设计的。这些将满足你更多的需求。最近，SAS和SPSS已经发展了包含其主要应用功能的PC版本。假如你购买了SAS或SPSS软件，就可以在你自己的PC上使用了。专家相信这些程序将导致数据分析的革命。许多大学买了上述产品的站点使用权，从而允许学生或教师在不同的计算机或实验室中使用。我们的经验是：这些程序是好用的，它包含很多“计算机运行程序”，仅通过点击就可以完成很多操作，这顺应了计算机工业的发展。不管你选择PC版的软件还是主机版的软件，数据分析都是行为实验最重要的部分之一（可以拨打312-329-3500获得SPSS软件的更多信息；SAS的可以拨打919-677-8000）。

推 荐 阅 读

我们希望，本章不仅提供了有关数据分析中高级统计技术的概览，也给出了该去何处获得有用信息的建议。对不同的统计方法和应用程序的有关细节，我们在此当然讨论得不够详尽。在本文中，我们提供给读者一些其他的资源，其中一些是高级的技术讨论，另外一些针对的是对基础概念的理解。一些资源可见Grimm和Yarnold（1995）、Pedhazur和SchMelkin（1991）、Stevens（1992）、Tabachnik和Fidell（1996），这些不是所有可用的参考文献，但它能让学生逐步熟悉不同的技术。

参 考 文 献

- Abelson, M. A. (1987). Examination of avoidable and unavoidable turnover. *Journal of Applied Psychology*, 72, 382-386.
- Anderson, J., & Gerbing, D. (1988). Structural equation modeling in practice: A review and recommended two-step practice. *Psychological Bulletin*, 103, 411-423.
- Bobko, P. (1990). Multivariate correlational analysis. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 1, pp. 637-686). Palo Alto, CA: Consulting Psychologists Press.
- Bray, J. H., & Maxwell, S. E. (1982). Analyzing and interpreting significant MANOVAs. *Review of Educational Research*, 52, 340-367.
- Bray, J. H., & Maxwell, S. E. (1985). *Multivariate analysis of variance*. Beverly Hills, CA: Sage.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Grimm, L. G., & Yarnold, P. R. (1995). *Reading and understanding multivariate statistics*. Washington, DC: American Psychological Association.
- James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal analysis: Assumptions, models, and data*. Beverly Hills, CA: Sage.
- Jöreskog, K., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Hillsdale, NJ: Lawrence Erlbaum.
- Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Levy, P. E. (1993). Self-appraisal and attributions: A test of a model. *Journal of Management*, 19, 51-62.
- Lipsey, M. W. (1990). *Design sensitivity: Statistical power for experimental research*. Newbury Park, CA: Sage.
- Long, J. S. (1983). *Covariance structure models: An introduction to LISREL*. Beverly Hills, CA: Sage.
- Pedhazur, E. (1982). *Multiple regression in behavioral research* (2nd ed.). New York: Holt, Rinehart & Winston.
- Pedhazur, E., & Schmelkin, L. (1991). *Measurement, design, and analysis*. Hillsdale, NJ: Lawrence Erlbaum.
- Schmidt, F. L., Hunter, J. E., & Urry, V. W. (1976). Statistical power in criterion-related validation studies. *Journal of Applied Psychology*, 61, 473-485.
- Spector, P. E. (1977). What to do with significant multivariate effects in multivariate analyses of variance. *Journal of Applied Psychology*, 62, 158-163.
- Stevens, J. (1992). *Applied multivariate statistics for the social sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York: HarperCollins.



第 18 章

进 行 元 分 析

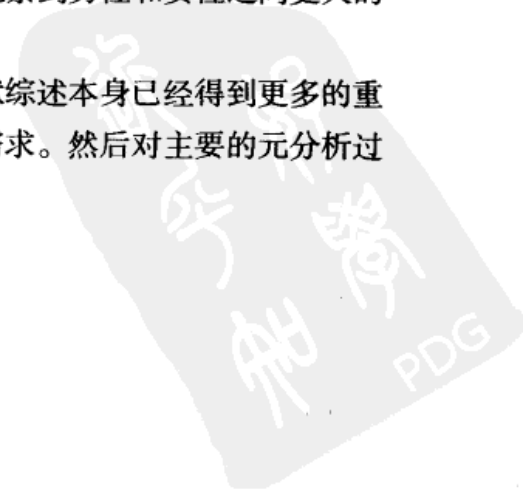
Harris Cooper

Nancy Dorr

文献综述常用于总结过去研究的结果，对过去研究发现中不一致的地方给出可能的解释，并指导将来研究的方向。一篇文献综述可以由两种方式来书写：传统的叙述方式、元分析方法。传统叙述性综述首先确认与感兴趣的主题相关的文章，然后检查每篇文章的结果是否支持某种假说，以得出全面的结论。结论通常以定性的方法得出，并且得出结论的过程通常取决于评论者个人。传统的叙述性综述由于在得出结论时很容易带入评论者的个人偏好，并且原有研究中的信息可能会被无意识地舍弃，因而常常受到批评（Cooper & Rosenthal, 1980）。元分析与传统的综述有着相同的目标，换句话说，元分析使用统计程序整合先前的研究结果，元分析过程就是一个整合相关研究结果和得出结论的专门过程。

Cooper和Rosenthal（1980）探讨了定量地（统计地）合并研究结果而不是定性地得出结论的意义。他们让大学教授和研究生来检验七项测试男性是否比女性更为固执的研究。参与者被随机地分配到以下两个组之一：（a）传统的陈述性综述条件，其中被试对7项研究结果做结论时，“被要求使用他们平时采用的任何标准”（p.443）；（b）统计的综述条件，其中教给被试如何用统计方法整合7个研究的结果。采用统计的评论者得出的结论更支持女性比男性固执，且比传统的陈述性评论者观察到男性和女性之间更大的差别。

随着元分析以及如何综合文献的精确标准的出现，文献综述本身已经得到更多的重视和投入。本章首先探讨元分析的历史，着重于它发展的需求。然后对主要的元分析过程进行描述。



元分析简史

1904年出现了我们今天称为元分析方法的最早应用。Karl Pearson (1904) 被要求根据11项相关研究中收集的数据, 检验一种伤寒疫苗的效用。对于每一项研究, Pearson计算了一项新发展出的统计检验——相关系数。他在根据因变量的性质区分的两组研究间, 平均了治疗效果的测量。根据相关的平均值, Pearson断定其他的疫苗更有效。他以跨世纪的远见提出: “改进免疫血清和剂量方法——两者相互高度关联——应该得到重视 (Pearson, 1904, p.1245)。”

接下来的便是75年后。在麻省剑桥, Rosenthal和Rubin (1978) 进行了一项关于在实验室、教室以及工作场所中行为的人际期望效应的研究回顾。他们发现不是11项而是345项研究与他们的假设相符。科罗拉多的Boulder、Glass和Smith (1979) 对班级大小和学业成就之间的关系进行了总结。他们在来自大约90万名学生的数据基础上, 发现了725项而不是345项的相关估计。这些研究者还收集了心理治疗有效性的评估, 描述了833项对治疗的检验 (Smith & Glass, 1977)。在密歇根州立大学和美国人事管理办公室工作的Hunter、Schmidt以及Hunter (1979) 揭示了866项对黑人工人和白人工人雇用测试差别效度 (differential validity) 的相似性。

以上每个研究小组都得出了不可避免结论: 传统研究综述的时代已经结束。他们意识到, 20世纪中叶大量研究人员进入社会和行为科学领域, 产生了预期的效果。在某些研究主题上, 已经收集到数量庞大的经验证据。这些数据涉及人类为什么按照某种方式去行为、感受和思考, 以及心理的、社会的、教育的和药物的介入对人类的行为、感受与思考的改善作用。这些研究者得出结论: 传统的叙述性综述有非常严重的缺陷, 包括对研究的选择性收集、通过事后检验把一些研究排除在考虑之外, 以及超出人类思考能力的、极为复杂的统计结果 (这结果是由于合并统计结果的数学认知不够完善而导致的)。

这三个研究小组独立地重新发现了Pearson解决问题的办法, 其他的研究人员立即步其后尘。其中有Richard Light 和 David Pillemer (1984), 他们做了一篇关于社会政策领域中使用研究综述的文章。Hedges 和 Olkin (1985) 提供了统计证据把定量综合确定为统计科学内的一门专门学问。Cooper (1982, 1989) 提出, 可以把研究综合过程与原始数据收集一样看待。他提出了一个类似于研究的综合研究综述的5阶段模型 (见表18.1)。这些阶段包括系统地阐述或提出问题、数据收集或文献搜索、数据评估、分析和解释以及公开呈现。对每一个阶段, Cooper整理了研究问题、在综述中的主要功能以及在可能造成综合结论中变异的综合程序上的差别。此外, Cooper建议, 一个由Campbell和Stanley (1966) 提出的由Cook和Campbell (1979) 扩展的推论效度 (如内部效度或外

表18.1 综述研究的整体概念

阶段特征	研究阶段				出版
	提出问题	数据收集	数据评估	分析和解释	
提出的研究问题	综述应该包括哪些证据?	为找到相关证据应使用哪些程序?	综述应包含哪些已找到的证据?	应该采用哪些程序,以便将文献作为一个整体来做出推断?	在综述报告中应该包含哪些信息?
综述的主要功能	建构用于区分相关与无关研究的定义	决定潜在的相关研究的来源	运用标准来分离“有效”与“无效”研究	综合有效的被选取的研究	应用编辑标准,区分重要与不重要的信息
引起综述结论变化的程序差别	包含在操作定义中的差别	包含在信息来源之间的差别	质量标准上的差别	推理规则上的差别	刊物编辑评判标准上的差别
综述结论潜在无效性的来源	操作细节的差别				
	狭窄的概念可能会使得出的结论不够明确	可获得的研究可能与综述的目标总体有质的差异	非质量因素可能导致对研究信息分配的不适当的权重	从噪音中区别出模式的规则可能不当	对综述程序的省略可能会使结论不能被复制
	肤浅的操作性细节可能会掩盖相互作用的变量	可获得的研究中的取样个体与目标总体不同	研究报告中的省略可能会使结论不可信	基于综述的证据可能被用来推断因果关系	对综述发现和研究程序的省略可能会使结论没有说服力

部效度)威胁,这一概念可应用于对综合研究的评估。Cooper还陈述了采用综合研究方法的先决条件:研究数据的二次使用者必须像数据的原始收集者一样,对他们研究方法的有效性负责。

元分析的过程

检索文献和对研究进行编码

让我们回到构造现今研究综合的程序问题。假设我们进入Polly C. Analyst博士的办公室,她想就补偿教育是否对青少年自尊产生作用这一问题做一综述研究。而且,假设Analyst博士已经完成了对相关研究的检索。她的搜索策略包括电子数据库(如PsycINFO)、传统程序、论文参考资料、相关期刊、对政府机构的咨询以及给当前活跃研究者写信(见本书的第2章及第3章)。注意,在Polly对相关研究进行检索时,她试图找出发表的和未发表的所有报告。这是很重要的,因为虽然一个报告有许多原因没被发表,但是这个研究没有拒绝虚无假设可能就是未发表的直接原因(也就是说,这个研究并未在接受补偿教育和没有接受补偿教育的被试之间发现自尊在统计上的显著差异)。如果Polly在她的元分析中没有包含未发表的研究,与包含了未发表的研究相比,前者可能会断定补偿教育对自尊有较大的作用。这样,考虑未发表的研究结果就可对教育与自尊间的关系做出更准确的评价。

进一步假设Polly的研究报告总共讨论了20篇验证补偿教育增强自尊这一假说的相关研究。对这20项研究的结果进行粗略的分析,Analyst博士注意到,有一些研究表明,接受补偿教育的个体比那些没有接受补偿教育的个体有较高的自尊;一些研究表明是否接受补偿教育的个体之间自尊并没有差异;还有一些研究表明,接受补偿教育的个体比那些没有接受补偿教育的个体自尊低。于是她决定探索导致研究结果不同的潜在原因。Polly进一步注意到,20项研究使用了多种程序。她考虑这20项报告的哪些特点会影响研究结果。比如,她可能会想,是否补偿教育对于男性和女性自尊的影响不同,是否补偿教育对不同年龄的被试自尊的影响不同,以及是否教育项目的类型或长度会对自尊产生影响。这样Polly系统地从报告的方法和结果部分中摘录出她想要用在分析中的每一项信息。Polly摘录了20项研究中样本的个人信息,可以帮助她确定对报告中编码的准确度和可信度。

投票计数

Analyst博士的数据收集现在已经完成。接下来她需要决定,如何整合20项研究的结果以得出有关补偿教育有效性的全面性结论。首先,她可以在20项报告中拣选、隔离出那些呈现结果与她自己的立场不一致的研究,摒弃其中在研究方法论上有局限的研究

(如没有对被试的教育水平进行随机分配), 而将剩下的支持性研究作为代表真相的事实。这样一种综合研究并不少见。然而, 它的结论极有可能受到怀疑, 对科学讨论贡献很少。

作为可供选择的另一个程序, Polly可以分析每一项报告, 并把它们归于下面三组中的一组: 表明教育能在统计上显著提高自尊的研究, 表明教育在统计上有显著相反效果的研究, 无法拒绝教育对自尊没有显著影响的虚无假设的研究。然后她可以宣布得票最多的那一组为赢。

这种投票计数的策略直观上很有诱惑力并且经常被使用。然而, 这种策略具有难以接受的保守性。在研究中, 一个研究的结果如果在统计上是显著的, 即当条件间的差异为随机的可能在5%或更小的水平上时(即 $p < .05$, 见本书第15章), 通常它就会被接受。在Analyst博士的评论中, 如果虚无假设是正确的, 即补救教育对于自尊没有作用, 那在20项研究中, 在偶然的因素下允许有一项(5%)研究得出教育是有用的结果。然而, 通过把20项研究分成三组, 投票记数策略则要求至少34%的研究结果得出有效的结果才能判定教育的有效性, 否则依旧看做偶然结果, 坚持虚无假设。因此这种投票计数策略可能, 且经常会导致放弃有效的研究。此外, 这种投票计数策略并没有在样本大小的基础上赋予研究差别的权重: 一个包含100名青少年的研究和一个包含1000名青少年的研究被给予相同的权重。由于大的样本更可能估计出准确的人口价值, 这是一个潜在的问题。因此从大样本中得出的效应应给予更大的权重。投票计数策略的另一个问题是, 每个研究中处理所带来的作用: 不管研究中一组接受教育的个体与一组没有接受教育的个体在得分平均值上差异有多大, 这些差异都被赋予了相同的权重。由于这些原因, Polly可以在她的元分析中报告三类研究中每一类所包含的研究数目, 但是她决定不能只依靠这些数据来得出结论。

合并概率

考虑到这些不足之处, Polly准备对每项研究中结果的精确概率进行整合。20项研究中的每一项都进行统计分析并检验联合概率, 对虚无假设(即接受补偿教育的青少年和没有接受补偿教育的青少年在自尊上没有不同)进行检验。为了合并每一项研究结果的概率, Analyst博士要找出对虚无假设的每一次测试, 确定与之相关联的概率。然后她采用几项统计公式中的一项, 得出在一系列研究中获得此结果的概率, 以之验证虚无假设是否成立。例如, 假定教育对自尊没有影响, Polly的结果可能表明, 20项研究得到此结果的合并概率是 $p < 0.03$ 。这样, 她就可以得出结论: 教育与高自尊显著相关。或者, 如果20项研究得到此结果的合并概率是 $p < 0.19$, 那么Polly可能会断定, 并没有明显的证据表明教育与自尊关联。

合并概率程序可以克服投票计数的不当权重问题。然而, 它自己也有着严重的局

限。首先，投票计数程序的过分保守，但合并概率程序则过分有力。事实上，对于已经产生了大量研究的实验处理，拒绝虚无假设很容易，以至于检验它变成了一项没有意义的工作。这一程序很少出现支持虚无假设、断定处理无作用的情况。对于Polly，这意味着，如果这是她使用的唯一的元分析程序，她很可能会得出教育与自尊相联系的结论。

效应值估计

合并概率阐述了一个效应是否存在的问题；它并没有给出一个效应是大还是小，是重要的还是无关轻重的。因此Polly认为，“补偿教育影响自尊吗？”这并不是她真正需要回答的问题。相反，她把问题重新设定为“补偿教育对自尊的影响有多大”？Polly的回答可能是零，也可能是一个正值或负值。正值一般表明效应值与元分析的假设是一致的。例如，如果Polly得出的效应大小是正值，这意味着受教育的个体比不受教育的个体有更高的自尊。同样，如果她得出的效应值是负值，这可能意味着，受教育的个体要比没有接受教育的个体有更低的自尊。然后Analyst博士可以正确地利用这一效应的大小来评估教育的支出收益是否平衡。而且，她可以提出问题：“什么因素影响教育的作用？”她意识到，对这一问题的回答可能让她对如何提高和引导教育干预提出更好的建议。

为了回答她的新问题，Polly决定在计算平均效应值的基础上得出结论，就像Pearson那样。由于她的研究涉及比较处理组和控制组——也就是接受补偿教育的青少年与没有接受补偿教育的青少年，所以她能通过计算 d 指标或标准平均差的方法描述效应值。 d 指标方法是一种没有刻度的、对两组平均数分离状况的测量。它以标准差为函数，描述了两组平均数之间的差异。图18.1描述了Polly综述中三种可能存在的结果。在图18.1A中，教育对于青少年报告的自尊没有作用，这样 $d=0$ 。在图18.1B中，接受教育的青少年的自尊分数比没有接受教育的青少年的自尊分数高4/10个标准差，这时 $d=0.40$ 。在图18.1C中， $d=0.85$ ，表明两组平均数间有更大的分离。

假定Polly的研究并没有比较两组的平均数，而是探索了两个连续变量之间的关系（例如，假设Polly想观察外向性与自尊之间的关系）。她不用计算 d 指标，只需使用Pearson的积差相关系数（即 r ，见本书中第16章）。与 d 指标不同， r 指标常在最初的研究报告中报告，获得效应值时并不需要额外的计算。

Polly在为每一项研究结果中计算了 d 指标之后，就可以平均所有的 d 指标，通过样本大小来赋予 d 的权重。这可能会告诉她对所有的20项研究，教育和控制条件之间存在多大的差别。这一平均的效应值忽略了研究的一些特点，比如样本的性别组成和教育时间长短。

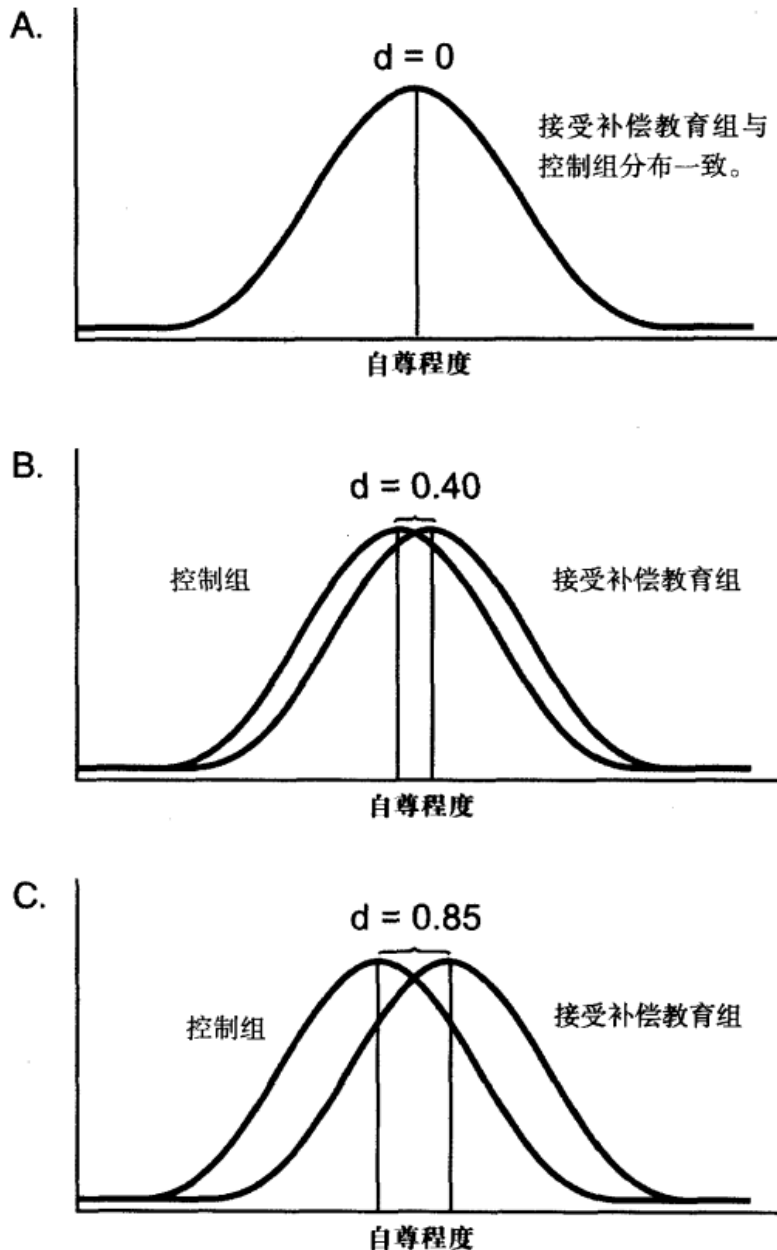


图18.1 自尊研究中接受补偿教育青少年与未接受补偿教育青少年的三种假设关系

效应值的影响

Polly接下来可能决定根据数据的不同特性，对教育与自尊的关系作不同的评估。当Polly浏览研究的结果时她注意到存在不一致。她可以检验，某些不一致是否由原来研究的不同特点和程序引起的。她对具有共同特点的研究分别计算了平均 d 指标。例如，她可能打算比较不同教育模式的平均效应值，区别使用大组授课的处理组和使用小组讨论的处理组。Polly可能也想看看，教育是否对不同类型的青少年有不同的效果，比如男性

和女性。她将研究问题扩展到探讨与教育效果大小相关的因素，这就显示出元分析的一个主要功能。说得具体一点，即使没有单个的研究对不同教育模式或青少年类型做出比较，通过比较不同研究间的结果，Polly的综合也可以为这些因素是否重要提供最初的暗示。

接下来，Polly可能会在统计上检验，这些因素是否与不同效应量（也就是不同的平均 d 指标）具有可信的相关——（Cooper & Hedges, 1994 ; Hedges & Olkin, 1985）这有时叫做同质分析，也可能是Polly综合分析中最有趣的结果。由于效应值是不精确的，即使都估计同一总值，它们也会产生变异。同质分析允许Polly检验，是取样误差本身引起这一变异，还是研究的某些特点、样本、处理设计或对结果的测量起到了作用。Polly按照潜在的重要特征把研究分组，再检验组间的差异。如果她发现存在差异，那么与这些差异相应的平均效应值就会有助于她对教育程序是否应该继续、为谁、何时以及多长时间等提出建议。例如，Polly 可以进行同质分析，来检验一个月的教育程序是否比一周教育的程序对自尊产生更大的影响。

如果同质分析是显著的，那么Polly可能会提议，一个月与一周的教育程序彼此之间效应值的不同并不是因为样本误差本身。Polly可能会计算出，一个月教育程序的平均效应值高于（根据样本大小加权）一周教育程序的平均效应值（如分别为 $d=0.55$ ， $d=0.05$ ）。这可能表明教育程序的长度与自尊相关，较长的教育程序可能受益更大。然而，Polly的同质分析并不允许她做出因果关系的推断。也就是说，Polly不能得出较长的教育程序引起较高自尊这一结论。这与随机分配有关：一项元分析不可能随机地分配各个研究。在一项研究中，除非研究者把被试随机分配到不同的条件中去（例如，一周的教育程序和一月的教育程序），否则就不能评估因果关系。不随机分配，中介变量作为真正原因的可能性就不能被排除。但不管怎么说，Analyst博士的同质分析在总结过去的研究、指导将来的研究中仍很有益处，尤其是如果原来的20项研究都没有检验教育程序长度的影响时。

总 结

在描述Analyst博士对完美的研究综合的追求时，我们的轻描淡写使这项工作显得很简单。Polly可能还会遇到缺失信息、整理混乱、数据值之间相关以及许多其他的问题。她如何处理这些问题，直接影响研究结果的可信性以及其他人对研究价值的评价。然而，如果社会科学研究的目的是对理性的决策有所贡献，那么研究综合就是我们方法论宝库中一个最为关键的成分，它与最初的研究一样需要严格和系统的标准。元分析方法可以帮助我们达到这些标准。

推 荐 阅 读

如果想了解关于元分析的更多信息,我们推荐Cooper的《整合的研究:文献评论指导》(*Integrating Research: A Guide for Literature Reviews*, 1989)以及Rosenthal的《用于社会研究的元分析过程》(*Meta-Analytic Procedures for Social Research*, 1984)。想了解关于元分析的最新内容,我们推荐参考Cooper和Hedge的《研究综合手册》(*Handbook of Research Synthesis*, 1994)。

参 考 文 献

- Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field setting*. Chicago: Rand McNally.
- Cooper, H. (1982). Scientific guidelines for conducting integrative research reviews. *Review of Educational Research*, 52, 291-302.
- Cooper, H. (1989). *Integrating research: A guide for literature reviews* (2nd ed.). Newbury Park, CA: Sage.
- Cooper, H., & Hedges, L. V. (1994). *The handbook of research synthesis*. New York: Russell Sage.
- Cooper, H. M., & Rosenthal, R. (1980). Statistical versus traditional procedures for summarizing research findings. *Psychological Bulletin*, 87, 442-449.
- Glass, G. V., & Smith, M. L. (1979). Meta-analysis of research on class size and achievement. *Educational Evaluation and Policy Analysis*, 1, 2-16.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Hunter, J. E., Schmidt, F. L., & Hunter, R. (1979). Differential validity of employment tests by race: A comprehensive review and analysis. *Psychological Bulletin*, 86, 721-735.
- Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.
- Pearson, K. (1904). Report on certain enteric fever inoculation statistics. *British Medical Journal*, 2, 1243-1246.
- Rosenthal, R. (1984). *Meta-analytic procedures for social research*. Beverly Hills, CA: Sage.
- Rosenthal, R., & Rubin, D. B. (1978). Interpersonal expectancy effects: The first 345 studies. *Behavioral and Brain Sciences*, 3, 377-386.
- Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, 32, 752-760.



第 19 章

测验的编制

Nancy E. Betz

本章中，我们首先讨论，为什么一个心理学家或研究者可能要决定编制一项新的测验或测量，而不是使用许多早就存在的测验或测量。然后，本章会对测验编制的主要方法之一——“基于构念”的方法进行详细的讨论。这一方法已经被广泛应用，并且如果应用合理的话，可以得出高质量的测验。在后面的部分，我还会提到测验编制的其他有用方法，但只是大略地提到。本章结束处有“推荐阅读”，可以提供关于这些方法进一步的说明。

为什么要编制一个测验

在做研究的时候，心理学家或研究者经常需要测量某些不可观察、但很重要的特点或属性。例如，一项关于顺序等级（出生次序）与智力之间关系的研究可能需要对智力进行测量，但智力并不是可以直接观测到的构念。其他主要的属性，包括能力、人格特质以及态度等，也是假定的存在。也就是说，我们不能实际地看到或接触到它们。一个人的智商既不标在人的前额上，也不印在一个人的社会安全卡上。换句话说，我们不能直接观测到智力。然而，我们可以发展出一项间接反映一个人智商的测验。

由于我们的研究思想通常建立在这一研究领域中先前研究的基础之上，所以先前研究中可能已对这些概念进行了测量。假定这些测量已经表现出足够的效力（见下面），使用已有的测量而不是发展一个新的测量似乎更加合适。这是因为，每一项研究都围绕一个给定的心理测验来积累知识，使得随后的研究更容易进行。总的来说，在一系列研究中使用高质量的测量，那么对理论和一般知识做出贡献就容易了。而且，几项研究证明，如果一项工具已拥有足够的信度和效度水平，那么新的测量只会浪费研究者的时间和精力。只有在需要测验提出的新理论或构念，或者在文献中对于一个给定的构念没有高质量的测量时，研究者们才应该发展他或她自己的测量。

为了确信一个有用的测量之前并不存在,研究者应该仔细回顾测验《词典》和《百科全书》,比如《心理测量年鉴系列丛书》(Kramer & Conoley, 1992) Beere (1990a, 1990b) 的与性别角色和妇女问题相关的测量摘要,以及Robinson, Slaver和Wrightsmann的《人格和社会心理态度测量》(*Measures of Personality and Social Psychological Attitudes*, 1991)。

评估测验质量

正如上面提到的,一般情况下我们推荐使用一项已经存在的测验或者高质量的量表。心理测验和测量的质量标准是强迫性的,在《教育学和心理学测验标准》(*Standards for Educational and Psychological Testing*)一书中有要求和详细的总结,该书由美国教育研究学会(AERA)、美国心理学会(APA)以及教育测量国家委员会(NCME; AERA, 1985)的联合委员会制定的。因为在本书的第6章中可以找到对质量评价的方法的详细讨论,所以这里我们只简要地回顾一下。

因为我们要发展测验作为假定构念的标示物,所以我们需要证明测验将能够测量出我们想要测量的东西。评估我们研究心理测量的质量标准,最重要的是信度和效度。信度是指我们可以以一种一致的、重复的方法来测量某些属性。内在一致性信度是指测量同一事物的项目之间内容是否一致的程度(即两个测验的项目或者一个测验的两种形式)。复测信度是指测量在时间上的一致性(比如,在最开始测验的两周后再重新测验)。

效度是指测验能否测量我们想要测量的事物的程度。有许多种效度(更全面的回顾见Walsh & Betz, 1995),但是主要的类型有内容效度、关联标准效度以及结构效度。内容效度是指测量某一特性的时候,选取的一系列代表行为是否能够反映该特性的整个行为范畴。也就是说,我们对一个人在多种情境下,或在一系列可能的“项目”上表现出来的行为感兴趣。虽然我们感兴趣,并且希望概括出整个领域的行为,但是在测量中很难包含所有可能与那一领域相关的情境。内容效度使我们能判断测验的内容是否代表了想要测量的内容总体。

术语关联标准效度是指对一个属性的测量在多大程度上与该属性的其他独立、外部的测量相关。这一外部标示叫做效标,可以代表我们实际感兴趣的行为。我们常使用测验分数或者其他测量来预测效标。例如,可用学业能向测验来预测学习成绩(平均学分绩点, GPA)。在学校中的成绩是人们感兴趣的东西,而测验分数与GPA之间的相关大小,就是测验在应用中的有效性的一个重要指标。

虽然内容效度和关联标准效度对回答“测验是否测量了它想要测量的事物”很有用,但是最直接回答这一问题的效度是构念效度。构念效度存在于与我们感兴趣的、与

构念相关的理论情境或假说内。理论上的构念通过测验或评估变为可以观察到的或者可操作的变量，然后通过研究那些测验评估之间的关系，就可以研究变量间假设的关系。一旦构念被操作化或变为可测量的，对它们之间关系的研究就可以开始了。在前面呈现的例子中，我们已经假定智力应该与学校成就和学习能力相关。由于我们明确了智力的概念，所以，我们也可以假定智力不应该与一个人的身高、眼睛的颜色或者出生地有关。对我们的智力测验得分与其他变量间关系的检验就是我们评估智力测验构念效度的手段。只有当观测到的变量间关系的模式与基于我们对构念理解的预测相符合时，我们才能说这个测验的构念效度比较高。虽然一个测验的信度和效度在测验编制之后得到评估，但是保证一个高质量测验的最好方法（可能是唯一的方法）是采用合适且仔细的测验编制方法。如果一开始就很好地编制一个测验，那么其可信性和有效性就会很高。高质量的测验编制是本章的核心。

测验编制程序

在开始之前，我们需要定义两个术语。首先，严格地说，测验一词是指一个事先有正确和不正确答案的测量，通常是一项能力或成就测验。例如，几乎没有人不同意问题“2+2等于几”的正确答案是4。在一项心理学入门课的考试中，问题“哪一位心理学家与最早的操作条件化的研究有关”的正确答案是B. F. Skinner。与测验相比，一个量表、调查或者问卷，并没有正确的答案，只是对要测量的刺激（或人）的描述。这些包括人格和职业兴趣调查以及态度、偏好和价值的测量。测验和量表在评估上还有差异，这些差异在后面的讨论中会有所涉及。然而，除非特别指明，下面测验编制的过程既适用于测验也适用于量表。

一般来说，测验编制按照下面的步骤进行：

1. 仔细地定义要测量的属性或构念；
2. 发展一系列符合在逻辑上与感兴趣特质相关的项目；
3. 在较大样本的被试身上施测这些项目；
4. 通过项目分析以及专家判断，进一步加工原有的项目；
5. 把修改后的测验施用于新样本的被试身上；
6. 在新样本的基础上，进行信度和效度检验，对数据进行标准化。

下面依次对每一个步骤进行讨论：

构念的定义

测验编制始于需要对测量的属性进行认真、详细的定义。这一定义应该包括属性是

什么以及不是什么；它应该包括对包含和不包含在定义中的行为和任务的详述。

如果我们为心理学入门课程编制一个期中的考试测验，首先要定义书的哪些章节以及讲座是考试要覆盖的，并指定不同内容相应的权重。但是定义一项人格或态度测量的内容、范围就没有那么简单。首先要仔细思考我们定义的属性代表的是什么。例如，这里是对构念无条件自重的定义，这个构思由Betz、Serling、Wohlgemuth、Harshbarger以及Klein（1995）提出，代表的是一种类型的自尊：

无条件自重是指个体把自己看做是一个有价值的人的程度，而且采用自身内部的评估标准，即一个人对自身的评价和接受是无条件的。换句话说，虽然个体可能会努力提高自身在一个或多个行为领域（如田径、学业、社会）里的表现，但他喜欢自己、接纳自己的程度并不取决于自己的表现。

我们可以编制一些反映这一定义的测验项目，例如，“我喜欢我这个样子。”“即使我犯错误，我仍然感觉我这个人很好。”

编制测验项目

心理学中测验和量表的编制应该包含许多项目。使用多项目量表有如下几个原因。首先，单个项目通常有相当大的特异性：一个孩子是否能拼写出裁判或汽化器一词可能取决于特定的知识背景或与其性别相关联。如果研究生入学考试（GRE）只用一个项目来决定是否被研究生院接受，那么大多数申请研究生院的人都会反对。第二，一个项目只能以很有限的方式对人做出区分。例如，“真实/虚假”或者“正确/错误”只能区分出两个水平，并且一般使用的五点Likert量表能够区分五个水平。然而，如果我们使用六个五点Likert量表，那么我们的分数等级范围为从1~30（ 5×6 ），可以允许区分30个水平。

第三，我们不能计算只有一个项目的量表的内在一致性信度。相对有限的分数范围会降低与其他变量的相关，这样就会降低效度系数。

一旦感兴趣的构念或维度被定义以后，那么测验的编制者就会发展出与维度的内容（即定义）相关的测验项目。这种项目发展要求测验编制仔细检查构念的定义，然后推论出能反映这一定义成分的特定行为或反应。

在编制测验项目时，我们一定要决定项目的内容和获得被试反应的方法。在能力和成就测验中，尽管其他的一些方法如真假判断和匹配也常使用，但是最常使用的是多重选择。量表，即没有正确或错误评判答案的测量，有许多获得反应的方法。最常用的就是累积等级或Likert量表（见Spector，1992），它要求应答者在一个连续的反应系列中，选择一个能最好反映自己当前状态的反应。在Likert量表中，最常使用的反应选择类型是同意、评估以及频率。同意常用一个从极同意到极不同意的五点或七点的连续变化体来标示。评估可以包括这样的连续体，如从根本没有描述到描述得很恰当，从一点也不确信到很确信。频率常使用这样的反应选项：从“极少”到“大部分”时间，或从“从

不”几乎“从不正确”到几乎“总是”或“总是正确”。

反应选择一定要使用数字，并且通常较大的数常代表着更多具有这一属性（如较高的自尊、较高水平的焦虑）。大约一半的项目应该采用消极词汇，以减少反应定势。当一个人计算整体分数的时候，消极的词汇项目要采用反向计分。例如，对陈述“我很少感到焦虑”同意或不同意的程度可以用Likert量表中的一个例子，人格特质“焦虑”的描述性也是如此。一般来说，在Likert量表中使用3~11种反应选择，但是大多数心理测量专家认为，5~7种是最优的。Likert量表常采用极赞同、赞同、中性、不赞同、极不赞同，分值从5~1（或者反过来）分别分配给五个反应选择。

测验项目应该与测验编制者所定义的构念具有合理的、逻辑的联系。测验编制者应该发展（自己编写或从可以利用的先前测验或其他材料中改编）一系列项目，数目要大于测验最后定稿所需要的项目数。大量的项目是必不可少的，这样，那些与感兴趣的维度不相关的项目就可以被删除。换句话说，虽然对于项目编制者来说，一个项目可能与维度有逻辑相关，但下面要讨论的项目分析过程就可能表明，它其实和该维度没有相关。

在发展样本上施行测验项目

测验编制的下一个阶段需要在一个初步的被试样本（通常称为发展样本）上试行测验项目。这一组中的被试应该可以代表测验本身想要施测的被试总体。换句话说，如果一个测验被设计用来测量7年级~9年级学生的学业成就，那么被试就应该是7年级~9年级学生的代表。同样，如果测验被设计用来测量成人的人格特质，那么就应该使用一个成人的代表样本。发展样本中的被试应该与最终测验处于同一条件下，从这一预测验中获得的反应将用于下面要讨论的项目分析。

通过项目分析和专家判断，进一步加工项目

无论是对构念的定义，还是对相关测验项目的发展，都要确保这些项目能够有效地代表这一构念。但在编制项目时，还需要使用有关项目其他属性的知识来进一步加工这些项目。在发展样本中实际地施测项目之后，我们需要清除那些并没有我们所希望的属性的项目，或者反过来说，选择那些包含特定和想要的属性的项目。在能力和成就测验中，两个最重要的属性就是项目难度和项目区分度。

项目难度

项目难度这一概念最开始被用于那些只有正确或错误答案的测验项目，如在能力和成就评估中。在这种项目中，项目难度通常按统计学，被定义为对一个项目正确反应的人数百分比。正确回答一个项目的人数百分比越高，那么这个项目就越容易，反之亦

然。项目难度水平通常用 p 值来表示：一个项目如果有70%的被试回答正确，那么 p 值为0.70，而一个项目如果只有20%的被试回答正确，那么 p 值即为0.20（某些计算项目难度的公式包括对猜测修正，见Nunnally & Bernstein, 1993）。

在 p 值基础上选择测验项目的方法相当大程度上取决于测验编制的目的，但是测验的难度直接取决于项目的难度。如果我们只包括非常难的项目，那么这个测验就会很难。同样，如果我们包括的所有项目都相对简单，那么这个测验也会很容易。

我们选择测验项目，以使难度的总体水平对于被试群体来说比较合适，测验产生分布相对较广的分数范围，更容易出现正态分布。如果一个测验没有得出一定范围的分数，那么它就不能为我们提供属性或特质上个体差异的信息。

为了阐述项目的难度如何影响测验结果分数，让我们首先考虑标准化或常态的样本中每个人都能正确回答的项目。项目的难度或 p 值是1.0，也就是100%的人回答正确。如果一个项目对于群体来说太容易，无法区分出群体中的个体，那它就不能提供关于个体的任何新信息。一个使用太多简单项目编制出来的测验可能会导致分数的偏态分布，产生“天花板效应”，即难以区别在分数分布上端的个体。

另一方面，我们假定有一个项目，常态样本中的每一个人都回答错误；这个项目的难度即 p 值为0。这一项目对于这一组人来说太难，与太容易的项目一样，它并不能提供给我们关于个体差异的新信息。由太难的项目编制出来的测验也会产生分数的偏态分布，通常被叫做“地板效应”，意味着它很难区分处于分数分布下端的个体。

一般来说，一组测验项目应该包括一系列的题目，难度多样化，比如 p 值分布从0.10到0.90。这就能产生一个难度合适的测验，可以有一系列分数值。这种测验项目的选择也确保了即使是最没有能力的人也能正确回答一些项目。

项目区分度

项目区分度这一属性对于绝大多数评估单一属性的测验来说都是很重要的。项目区分度是指人们对一个构念的某个项目的反应得分与在整个测量上得分之间的相关程度。如果项目反应与整个量表上的操作没有关系，那么我们就有理由猜测，这个项目并没有测量出它想要测量的东西。例如，假定一个班级中最聪明的孩子在测验中得99分，而他回答错误那道题却被最差的学生（总共只做对了10个）答对了，这样我们就会怀疑，这个题目是否真的能测量其学习成绩。或者假设其中的一个项目是我们用来测量焦虑的，但在整个焦虑测验中得分最高的（最焦虑的）人对那个项目的反应却是否定的：同样我们会怀疑这一项目作为焦虑测量的有效性。测验项目的辨别能力与其内在一致性信度直接相关。可靠的测验需要区分度高的题目。

区分度很差的项目可能是因为表达不够明确、在意思上模糊，或者与感兴趣的维度毫无关系。项目区分度的标准允许测验建构者选择“好的”项目（即那些实际上能测量

出感兴趣维度的项目)，把“不好的”项目从测验中排除掉。

项目区分度有许多标准。在辨别力的测验中，广泛使用的两个相关系数是二列相关系数和点二列相关系数。一方面，点二列相关系数 r_{pb} 是Pearson积差相关系数，描述了一个二元变量（如真实/虚假、正确/错误、男性/女性）和一个连续变量（比如一个测验上的总得分）间的关系。另一方面，二列相关系数 r_{bis} 描述了二分变量（一个假定代表正态分布的二分变量）和连续变量间的关系。 r_{pb} 和 r_{bis} 的计算是根据通过和没通过某项目的两组人的平均测验分数比较而来的，与测验项目本身的难度有关（想了解更多关于这些以及其他相关系数的信息，请参考Glass & Hopkins, 1984）。具有足够区分度的项目必须保证，对该项目回答正确的所有人的测验分数都要高于对该项目回答错误的人的平均测验分数。

对与项目难度相关的能力测验，一个好的项目选择策略就是在每个难度的范围内选择 n 个辨别力最高的项目。假定我们有10个难度范围（ $p=0.01 \sim 0.10$, $p=0.11 \sim 0.20$, …… $p=0.90 \sim 0.99$ ），并且想要一个包含100个项目的测验，我们可能会使用每个水平上10个最有辨别力的项目。

为量表选择项目

对于非能力测验，决定某个项目去留的最重要标准是其区分度，它通过项目与总分的相关度（通常是Pearson相关系数）来考查。让我们回到前面讨论的对无条件自重的测量。被试在一个五点的Likert量表上进行选择反应：极赞同（5），赞同（4），不确定（3），不赞同（2），极不赞同（1）。如果极赞同反映的是高自尊，那么我们会希望做这一反应的人在整个量表上的总分（把量表中所有项目得分相加）也高。这样我们可以在项目“分数”（1~5）和总分之间计算相关值。一般认为，项目与总分相关系数低于0.30（或者是低于0.20）时，表明该项目辨别力差。

选择项目的其他标准包括项目的平均“分数”（类似于项目难度）、对项目反应的变化度（如果每个人对一个项目的反应为“极赞同”，那么这个项目就不能反映出个体差异）以及与“社会赞许性”测量之间的相关（最好是较低）。

在量表编制中因素分析的使用

当我们想观察一组变量的根本结构或基本维度时，当我们想把变量集缩小时，我们可以使用因素分析。例如，对人格感兴趣的研究者可能想了解，20或30个对人格不同侧面的测量可以归结成更为基本的维度，或者假定一个公司使用20个不同的测验来选拔担任某种工作的雇员；因素分析可以用于寻找这20个测验所反映的可能是4个或5个可能的基本维度。通过对这4个或5个基本维度的测验，可以做出与原有的20个测验一样准确、有效的预测。

在测验编制中，因素分析常以两种方式使用。一方面，如果一个构念被假定为是单维度的，那么因素分析应该得到一个且只有一个维度，或用因素分析的语言来说，得到一个一般性因素，尽管它可能同时伴随一个或多个小因素。另一方面，一个研究者可能希望考察一个已存量表或新量表的基本维度。这时，因素分析也是一个有用的方法。在某些情况下，因素分析可用于对一个量表的亚量表得分的计算，从而得到关于被测试个体更多的信息。例如，在一个初级数学知识测验中分别计算乘法、除法、分数运算的得分。

因素分析从计算量表所有变量间的相关——相关矩阵——开始。因素分析的结果是因子矩阵，它显示了重要的基本因子的数目以及因子中每个原始变量的负荷（权重）。因子负荷就像相关系数：它越大（无论是正向的还是负向的），那么变量和因子间的关系就越强。用因素分析的语言来说，负荷也可以看做是与由基本因子所解释的变量方差的多少有关。因子负荷的平方就是变量和因子间的方差比率。

表19.1给出了一个对包含六种类型项目的智力测验进行因素分析得出的假想因子矩阵。前三个测验（词汇、言语类比以及阅读理解）在因子Ⅰ上的负荷很大，我们可以把这些测验测量出的根本能力叫做“言语能力”。后面三个测验（加法、乘法以及分数运算）在因子Ⅱ上的负荷很大，它们可以把这个因子叫做“运用算术的能力”。要想了解关于因素分析的更多内容，可以看本书的第17章以及Tinsley和Tinsley（1987）。

表19.1 根据六项不同任务上得分分析得出的假想因子矩阵

测验	因子Ⅰ	因子Ⅱ
词汇	<u>0.90</u>	0.40
口语类比	<u>0.80</u>	0.30
阅读理解	<u>0.75</u>	0.35
加法	0.30	<u>0.80</u>
乘法	0.35	<u>0.85</u>
分数运算	0.77	<u>0.70</u>

注意：对一个因子的负荷很大的变量下面被画线。

量表编制的其他方法

因本章的篇幅有限，关于测验或量表编制的其他几种方法这里就不详细讨论了。接下来我们对这些方法作一简要介绍，并为对此感兴趣的读者提供一些信息来源。一般来说，心理测量和测验的教科书会对这些方面详加阐述，可参考Nunnally、Bernstein（1993）以及Walsh和Betz（1995）的著作。

经验量表编制

与量表编制的理性方法相比，经验性量表的编制是基于人们在某个属性的反应差别，而不是基于项目内容本身。换句话说，测验编制首先搜索那些能预测测量效标或状态的项目。例如，明尼苏达多相人格问卷（MMPI）原来是为临床工作者检查那些遭受严重心理失常（如精神分裂症）的人而设计的；把已经诊断为患有精神分裂症的人在大量

项目上的反应与那些“正常的”个体的反应进行比较——在这个研究中，正常个体是指到精神分裂症病人所在的医院来访问病人的人。那些病人组的反应与正常个体的反应有显著差异的项目就被用于编制MMPI的精神分裂症量表。其他经验性测验编制的例子有Strong兴趣问卷(SII)中的职业量表。在这个例子中，效标组是那些满意自身职业(如内科医生)的人。对某职业感兴趣的人与一般大众如在对某项目的反应上有显著差异，则该项目就被包括在职业测验中。

态度量表编制

态度可以被定义为一个人对给定的一类刺激的看法或反应。刺激包含的范围很广，如可以包括如下事物：对政治候选人、人种/种族群体、妇女在社会中的角色、宗教、男性同性恋的生活方式、对人工流产等的态度。态度量表编制的技术问题超出了本章的范围，感兴趣的读者可以查阅Dawis(1987)、Fitzgerald和Hubert(1987)以及Maranell(1974)。四种众所周知的方法是Thurstone、Likert(在前面已经提及)、Guttman量表和配对比较方法。也可参考Edward(1957)的经典著作《态度量表的编制方法》。

项目反应理论

前面大多数对于测验编制的讨论都是基于我们所谓的经典测验理论。经典测验理论假定每一个个体在测验中都有一个“真实的分数”，并假定测验分数代表了这个真实分数外加测量误差量。我们假定，在真实分数和测量到的分数之间存在着线性关系，信度系数表明了可被归为真实分数和观测分数方差的比率。

虽然经典测验理论今天仍然广泛应用，但测验理论的一个最新发展是潜伏特质理论、项目反应理论(IRT)或叫项目特征曲线理论(ICC)(更多的综述可见Weiss, 1983; Hambleton, 1989)。这一理论把有关项目难度和项目区分度的有关信息代入数学方程，以指导测验编制。

具体说来，潜伏特质理论(或者是IRT)基于测验分数反映个体拥有“潜伏特质”的量这一假设。这个潜伏特质，由希腊字母 θ 来表示，类似于构念效度中所讨论的假定的构念。另外，项目可以使用三个属性中的一个或多个来描述：项目难度、项目区分度以及多选测验中正确的猜测比率。这一理论提供了数学模型或方程，可以表达对一个测验项目的反应与一个人拥有的潜伏特质之间的关系。方程使用了关于项目难度、项目区分度、正确猜测比率的信息，来估计那些做出某一特定反应的个体所拥有的特质质量。

在测验编制中使用潜伏特质理论有许多优势。例如，每一个测验项目都可用来对个体能力进行估计；使用的测验项目越多，估计越精确。在实际中，这意味着不同的个体可以采用不同的项目或不同数目的项目进行测试，并测量到同一个“潜伏特质”，而且分数可处于同一数值量表上，有利于互相比较。

这一思想已经被用于能力的“适应”测验（Wainer, 1990; Weiss, 1974, 1982）上，其中被试接受一系列根据他们能力水平而“特制的”的一组项目。例如，在一个两阶段测验中（Betz & Weiss, 1973），初步的“路线分配测验”是由计算机管理和计分的，然后根据分数，给被测验者更难或更容易的第二阶段测验。这样，被测验者如果在路线安排测验中做得不好，与在测验中做得好的被试相比，他们在第二阶段中就更可能接受一些较容易的项目。适应测验程序可以使测量的信度最大化，同时使对测试一个被试所需的项目数量最小化。

总之，在测验编制中，虽然潜伏特质理论以及适应测验相对较新，但是他们代表了未来的发展趋势。一些主要的测验，如军队服务能力倾向测验（ASVAB）已经有了计算机管理的版本（Moreno, Wetzel, McBride, & Weiss, 1984）。

总 结

本章试图综述测验编制的基本原则，但限于篇幅，我们不能对所有主题内容进行详细讨论。我们建议，想编制测验或量表，以用于研究或应用目的的同学应该选修一些测量理论及方法的课程，或是查阅这里所提到的诸多文献。编制量表是一种使重要的心理学构念操作化的方式，但只有通过细致和明智的工作，它才能促进科学进步。

推 荐 阅 读

有关这一主题的进一步信息，感兴趣的读者可以参考：《教育学和心理学测验的标准》（*Standards for Educational and Psychological Testing*, AERA, 1985），Cronbach和Meehl（1955）的著作，Nunnally和Bernstein的《心理测量理论》（*Psychometric Theory*, 1993），Spector的《累积等级量表的编制》（*Summated Rating Scale Construction*, 1992）。

参 考 文 献

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: Author.
- Beere, C. (1990a). *Gender roles*. Westport, CT: Greenwood.
- Beere, C. (1990b). *Sex and gender issues*. Westport, CT: Greenwood.
- Betz, N. E., Serling, D., Wohlgemuth, E., Harshbarger, J., & Klein, K. (1995). Evaluation of a measure of self-esteem based on the concept of unconditional self-regard. *Journal of Counseling and Development*, 74, 76-83.
- Betz, N., & Weiss, D. J. (1973). *An empirical study of two-stage ability testing* (Research Rep. No. 73-4). Minneapolis: University of Minnesota, Psychometric Methods Program.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Dawis, R. V. (1987). Scale construction. *Journal of Counseling Psychology*, 34, 481-489.
- Edwards, A. L. (1957). *Techniques of attitude scale construction*. New York: Appleton-Century-Crofts.

- Fitzgerald, L. F., & Hubert, L. (1987). Multidimensional scaling: Some possibilities for counseling psychology. *Journal of Counseling Psychology*, 34, 469-480.
- Glass, G. V., & Hopkins, K. D. (1984). *Statistical methods in education and psychology*. Englewood Cliffs, NJ: Prentice Hall.
- Hambleton, R. K. (1989). Principles and selected applications of item response theory. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 147-200). New York: Macmillan.
- Kramer, J. J., & Conoley, J. C. (Eds.). (1992). *The eleventh mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.
- Maranell, G. (Ed.). (1974). *Scaling: A sourcebook for behavioral scientists*. Chicago: Aldine.
- Moreno, K. E., Wetzel, C. D., McBride, J. R., & Weiss, D. J. (1984). Relationship between corresponding ASVAB and computerized adaptive testing (CAT) subtests. *Applied Psychological Measurement*, 8, 155-163.
- Nunnally, J., & Bernstein, I. (1993). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Robinson, J. P., Shaver, P. R., & Wrightsman, L. S. (1991). *Measures of personality and social psychological attitudes*. San Diego: Academic Press.
- Spector, P. (1992). *Summated rating scale construction*. Newbury Park, CA: Sage.
- Tinsley, H. E. A., & Tinsley, D. J. (1987). Use of factor analysis in counseling psychology research. *Journal of Counseling Psychology*, 34, 414-424.
- Wainer, H. (Ed.). (1990). *Computerized adaptive testing: A primer*. Hillsdale, NJ: Lawrence Erlbaum.
- Walsh, W. B., & Betz, N. E. (1995). *Tests and assessment*. (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Weiss, D. J. (1974). *Strategies of adaptive ability measurement* (Research Rep. No. 74-5). Minneapolis: University of Minnesota, Psychometric Methods Program.
- Weiss, D. J. (1982). Improving measurement quality and efficiency with adaptive testing. *Applied Psychological Measurement*, 6, 473-492.
- Weiss, D. J. (Ed.). (1983). *Latent trait test theory and computerized adaptive testing*. New York: Academic Press.



第 20 章

使用档案数据

Barbara H. Zaitzow

Charles B. Fields

档案数据是什么

一个社会科学家通过对次级数据集的分析，经常能回答一些迫切的研究问题。次级数据是已出版的数据，收集这些数据是要“获取一般性的信息或是回答某个特定问题而设计的研究工作的一部分”（Stewart & Kamins, 1993, p.2）。使用档案数据是一种非互动的数据收集方法，因为研究者并不与被研究的被试本人接触，这样，研究中就可以消除由研究者的活动效应引起的潜在错误（Webb, Campbell, Schwartz, Sechrest, & Grove, 1981）。然而，只有经过最初的数据收集，次级的数据才能存在。这样，次级数据就会受到最初的观察者所使用的方法的影响，这是本章要阐述的几个问题之一。本章的目的是提供社会科学研究使用中档案数据集的概述。

Emile Durkheim是最早利用别人收集的信息的人之一，他是社会学的奠基人。19世纪末期，在题目《自杀》（1897/1966）的研究中，他使用来自许多国家的大约26 000项报告的记录，检验了农村与城市两种社会环境间自杀比率的变化。今天我们可以根据美国统计局出版的材料，来讨论人口增长和衰退的趋势，也可以使用联合国收集的统计数据，来研究世界经济变化和发展的模式。

我们可以研究小说、歌曲、电视节目，以及其他的大众媒体产品，来更好地理解我们的社会，也可以深入探究过去，使用社会所保留的记录或祖辈们的日记和信件，来研究今天的家庭和上个世纪的家庭有何差别。Terman在斯坦福大学进行了“天才研究”的突破性研究工作，他对1600名有天分的加州儿童从1922年开始追踪，直至如今。这个研究与加州大学Berkeley分校的人类发展研究所的几个著名的纵向研究合作，开创了数据档案的研究方法，这些数据档案使得研究者能够检验来自不同出生群体的人们生活的历史影响。

什么时候应该考虑使用档案数据

档案数据可通过多种方式应用，且可以应用于许多研究问题。累积的数据可以用来探究影响犯罪的社会因素。例如，为了研究犯罪与贫穷之间的关系，犯罪学家使用统计局收集的关于城市地区接受福利、单亲家庭的人数以及收入等数据，将这些信息与在同一地区的犯罪记录相联系（Blau & Blau, 1982; Messner, 1983; Sampson, 1985）。使用这两种档案数据源加深了我们对整个社会趋势与犯罪率模式关系的理解。档案数据也可用于比较10年间不同人种或种族群体的社会经济特点，探究他们在社会经济地位上的差异在增大还是在减小，或者检验自第二次世界大战以来工业活动的市郊化，并分析这一趋势对于中心城市少数民族失业群体的影响。

社会科学家们从经常性的民意测验中获得了缺失的数据。许多调查组织，如负责盖洛普民意测验的美国民意调查研究所，Roper组织和Louis Harris、Associates，在过去的几十年里曾负责数以百计的调查。类似的调查数据对于研究社会变化和习惯非常有用。各种调查提供的数据极为广泛，只要研究者想得到，他就有做次级分析的机会。

缺点

次级分析和调查分析非常相近（Goodard & Villanova, 见本书第7章）。调查研究和次级分析之间的基本区别在于，后者中的数据是由其他人为了其他目的而收集的。这意味着研究者对数据或者数据收集过程没有控制力。其导致的结果是，由于数据没有用合适的形式来收集，所以有时很难形成合适的操作定义。Emile Durkheim 19世纪关于自杀的研究表明了这一方法的某些危险。他依赖于官方的死亡记录来考察不同类型的人的自杀率，但是他没有办法来检查多少事故被错误地分类为自杀，或者多少自杀被记录为事故或其他原因致死。因此，研究者一定要谨防这些歪曲。

在最初的研究中，一些问题可能会以一种并不适于研究者目的的方式呈现，给被试或者作反应的被试可能不是当前研究者打算选择的被试。“在一个特定的研究中，一定要注意一个研究到底收集了哪些信息（Stewart & Kamins, 1993, p.23）。”在数据收集过程中对使用方法（如样本特点、反应比率、面试或者调查的原始记录、测量工具、代码形式等）的了解，可以保证次级数据使用所得结论的效度和概括性。

对社会科学研究来说，人口普查数据的主要缺点是社会科学家感兴趣的许多变量在计数调查中并没有报告出来。例如，没有收集有关情绪、态度、价值或信仰的信息。因此，对社会心理学研究感兴趣的研究者一定要在别处寻找次级的数据源，或他自己从事现场研究或实验。

“次级数据是旧数据”（Stewart & Kamins, 1993, p.6）。例如，人口普查数据发

表的耽搁在时间上可能会产生测量方法、结果获得以及结论推广方面的局限性。例如，如果你使用档案来研究美国婚前性行为的变化，那对结果的解释必须考虑到复杂的社会历史事件。很明显，对于美国的老年人来说，“直到结婚才发生性行为”的原因与当今大多数人的原因有质的差异，后者是担心“第一次可能招来艾滋病的死亡判决”。在两种情况下，反应都表达了禁欲的想法，然而原因却可能极不相同！重申一下，当一个人使用档案数据时，必须注意社会历史问题（Gergen, 1973）。很显然，这里存在权衡的问题；一方面要考虑数据的有用性；另一方面要考虑问题的难易程度。因此，研究者无论何时使其他人收集的信息，都要仔细地以一种适合于自己研究目的的方式小心谨慎地解释数据。

优点

次级分析使研究者从收集和编码数据所需的时间、花费和责任中解脱出来。档案数据的使用使社会科学家们能够用较低的预算进行一些研究，而且在美国的人口普查中取样和测量误差是所有同类调查数据源中最低的。但不管怎么做，人为和机械的误差还是会发生，降低了一些数据的准确性。用美国统计局的话来说（1964）：

这样的误差包括没有从应答者处获得必要的信息、获得不一致的信息、在错误的地点收集信息、信息记录不正确、历次数据输入的不一致、相互联系的项目间不一致。调查员没有遵照取样指导而产生取样偏差。工作人员在编码和整理中会出现错误，电子处理中也有错误发生。（p.lxxxv）

为了减少这些错误，人口普查局复查调查员要对手工编码和编辑进行检查，核对图表，利用统计方法（如样本数据的比率估计）来控制数据总体。通过这样的努力，出版的报告中的错误通常可以保持在一个可以接受的较低水平，对这些数据的次级分析一般不会产生误导的结果。

档案数据对于次级分析尤其有用的原因是，大多数变量不仅可以跨区域（在同一个时间点），而且可以跨时间（在不同的时间点）。这使得研究者能够估计特定的社会关系并检验它们的历史趋势。例如，犯罪学家们广泛利用官方关于美国各种犯罪频率的统计，尤其是由联邦调查局提供的“总体犯罪报告”（Uniform Crime Report），但他们还总得提防这些统计中的歪曲之处。例如，我们知道许多犯罪（如非法使用毒品）没有报告给警方。使这一问题更为复杂的是，报告的比率根据犯罪的类型变化而变化。凶杀很可能引起警方的注意，然而强奸经常不被报告。试图修正这一差异需要研究者收集自己的数据来检验官方统计的准确性。如果我们知道可用的数据的准确程度，那么它的价值显然就被提高了。而且，克服次级数据中出现的问题的最好办法是利用多种信息来源。

在一个更为实际的水平上，在你决定使用档案数据时，一定还要评估其他情况，包

括：（a）数据能被转移到你的计算机系统中吗？（b）数据的相关文件说明（你可能根本无法获得它）是否足够你对数据进行分析？（c）你是否有足够的专业知识和耐心来处理由不同集合水平、分类模式、分析软件所产生的问题？

Elder、Pavalko以及Clipp（1993）在关于档案数据的工作中，指出了如下常见问题：

1. 档案数据永远不会像一个人期望的那样精确。考虑到这一点，研究者必须面对挑战，在给定的时间和资源条件下把数据整理成需要的模式。
2. 手边的数据通常反映了原来调查者的视角，这些烙印反映在研究问题、数据收集过程以及分析方法之中。
3. 纵向数据档案并不保证能进行纵向数据分析。横断数据更为普遍，这些横断数据常常需要转换成纵向生活记录的形式。
4. ……（这些）研究可能利用了定量和定性数据，对两种数据有效使用需要细致的计划，这样才能用于分析相同的问题。
5. 使用档案数据的基础是这些数据的长处，我们不能通过废弃或忽略其不足来维护其合理性（p.11）。

仔细地考虑以上这些要点对于决定档案数据是否有利于你的特定研究是很有必要的。要想使得数据可用，你一定要清楚地知道你打交道的对象。

档案数据的获得

正如你可能意识到的，近几十年里，学院和大学的研究者、私人组织和公司以及政府机构已经做了成千上万的调查。大部分调查都形成了计算机化的数据档案，其中许多可以被研究者利用（对这些资源的回顾可见Stewart和Kamins，1993）。现有的数据主要有三种来源：（a）各种公共和私人数据档案；（b）美国人口普查及相关资源；（c）适合于内容分析的出版物或广播媒体。下面的讨论将集中在前两类，因为它们代表了“已经收集”的数据。

公共和私人数据档案

许多从事研究的学者愿意把他们的数据分享给那些与自己研究兴趣有轻微差别的人。事实上，一些人常常通过一些机构与别人分享数据。例如，杜德克利夫学院亨利·A·墨莱研究中心（Henry A. Murray Research Center of Radcliffe College）是“国家的心理学、精神病学、社会学、人类学、经济学、政治科学以及教育领域中的数据贮藏处”，它贮存了许多可用于次级分析的信息（Elder等人，1993，p.83）。同样，政治与社会大学研究院（ICPSR）也是一个国家性组织。

政治和社会研究校际联合会

它成立于1962年，作为密歇根大学调查研究中心与21个研究性大学的合作机构，ICPSR现在已经涵盖美国和加拿大的325所学院和大学以及世界上几百个研究所。ICPSR通过归档、接收、加工以及分发，从各种来源获得关于社会现象的计算机数据。从18世纪90年代起的美国大选的信息到最近的联合国会议材料，有关的数据都可以得到。

数据和支持性文件可以通过成员机构免费（或花费很少）获得；非成员也可以获得ICPSR数据，但是费用会贵一些。有四个“等级”的数据可以利用，这些数据按照ICPSR必要的加工程度和水平分类：

一级：数据已经被检查、被修正（如果必要的话），使用了ICPSR的特定格式。为了使数据的可使用性最大化，咨询过原研究者，ICPSR已经对它们进行了再编码和重组。通常可以得到一本可被计算机阅读的代码书。

二级：这些研究已经被检查且转换为ICPSR的特定格式。大多数的非数字编码已经被删除。当需要数据时，可提供文件（代码书）的副本。

三级：研究经ICPSR的工作人员检查过数据记录量以及数据存贮地（由调查者的代码书指明）。数据通常以最初研究者提供的格式呈现的。

四级：这些研究由ICPSR按原观察者提供的形式分发。提供的文件是按照原来收到的材料的副本（ICPSR, 1994, p.xxiii）。

为了有效地使用它们的数据集，ICPSR也提供了各种辅助与支持。例如，ICPSR通过书面材料、专门的部门、讨论会以及电话咨询等方式，提供会员设备和数据存储媒体的选择、合适软件的确认以及计算机网络技术使用方面的建议。ICPSR还有一个重要的目的就是促进使用新的“正在急剧改变社会科学教学和研究潜力”的技术（ICPSR, 1994, p.viii）。

一些心理学的学生可能比较感兴趣的数据集也可以从ICPSR获得，如：

- 美国人看自身的心理健康，1976（ICPSR 7948）——由国家心理健康研究所赞助的全国性调查数据。调查（2 264 850个变量的样本）集中于婚姻、父母的身份、受雇用情况、闲暇时间、成就和能力动机以及一般社会关系等。此外，还收集了对帮助的寻求、寻求专业心理帮助的意愿程度、心理辅导的安排机制以及对接受到的帮助的评估等信息。
- 人一生中的婚姻不稳定性（U. S.）：“三波小组研究”，1980—1988（ICPSR 9747）——这一数据收集试图确定整个生命过程中婚姻不稳定性的原因。1980年采用电话调查方式收集的全国55岁或低于55岁的已婚个体样本（ $N=2\,033$ ），然后在1983年和1988年再次进行调查。询问了经济来源、受雇用情况、子女和生活目标等问题，以确定可能引起离婚的原因。

- 父母与儿童调查, 1990 (ICPSR 9595) ——由国家儿童委员会实行。调查的目的是评估美国家庭的态度、福利以及生活状况等。样本由1 738名父母和929名10~17岁的孩子组成, 其中黑人和西班牙儿童的家庭取样居多。
- “天才”项目公共使用文件, 1960—1976 (ICPSR 7823) ——一个纵向调查, 用来评估促进或抑制人类才能发展的个人、教育因素。在1960年, 9~12年级的400 000名 (这里包括4 000个子样本) 学生接受调查, 内容涉及兴趣、计划、家庭背景、认知技能等。在1年、5年以及11年后进行关于高中毕业后教育、家庭发展、计划和抱负以及生活满意度的后续调查。

其他的政府文件

许多政府机构进行调查以及其他形式的研究, 然后出售计算机数据档案。FBI著名的“统一犯罪报告”(UCR) 就是这样一个的数据来源的例子。UCR是美国今天最常使用的官方数据源之一, 并且一直在指导犯罪干预和控制的政府决策方面发挥强有力的影响。然而, 在选择和使用UCR以及类似的数据库的时候, 必须考虑到某些限制, 并且在使用之前一定要仔细检查这些方面。例如, UCR一直基于当地警察部门的自愿报告。这样, 可以发现城市和乡村管辖区之间报告的差异较大, 后者报告较少。可能更严重的缺点是对犯罪受害者的调查。由司法统计局出版的“国家犯罪调查(NCS)”表明只有少于50%的严重犯罪会被报告给警方(Bastian & DeBerry, 1994)。因此, 建议使用多种数据源, 给研究的对象提供一个更准确的描述。

虽然人口统计局(在下面讨论)可能是档案数据的最佳来源, 但是从其他政府机构也可以得到很丰富的信息。在联邦、州、地区级别的政府机构、司法系统以及各种立法和行政管理机构, 不断地产生大量的信息。虽然由于某些原因这些信息中的大部分不能通过传统方法获得, 但1966年的信息自由法令为公众获得这类信息提供了诸多方便。

美国政府也有一个注册系统记录出生、死亡、结婚以及离婚事件。健康和人类服务部除出版每年的总结报告外, 还有一个“每月的重要统计报告”, 它是对这些主题感兴趣的研究者的一个重要资料来源。

美国人口统计数据

数据的一个主要来源是美国的人口普查, 它在一个较好的图书馆就可得到。它每10年进行一次的统计, 包括了小到个城市街区, 大到整个国家的人口和居住特点的全面数据。粗略浏览任何一本主要的社会学杂志都可以发现, 对社会学来说, 人口普查是数据最重要的来源之一。除了人口和居住情况以外, 对于制造业、批发贸易、零售贸易、服务业、农业、矿业以及政府等也有经常的统计。所有这些统计的结果都可以在任何一个较大的图书馆查阅到, 也可以在计算机磁带或光盘中找到, 这些磁带和光盘可以在人

口统计局或任何大学和私人公司处购买。

人口普查每10年进行一次。然而，人们还需要更多即时的数据。这一需要可以由“即时人口调查”（Current Population Survey）来完成。这一调查由“劳动统计局人口分局”持续进行，每年都会更新关于人口数量、年龄结构、种族构成、收入、教育以及雇用和失业等方面的信息。此外，在特定的领域内，还有常规的和专门的（如一次性的）补充调查。如果你需要关于这些主题的最新数据，就查一下“即时人口调查”。

ICPSR档案也包含许多可被研究者利用的人口普查数据集。可以获得1970年、1980年、1990年的完整普查数据，还有从原始统计问卷中得到的多种样本数据，也可以查到1860年（城市家庭样本）、1880年、1910年（黑人家庭的样本）而后一直到1960的不完全的普查数据。

需要提醒的是：虽然人口统计局和其他政府机构有独立于政治干预的传统，但各种管理部门能够决定报告的主题以及数据多久出版一次。例如里根政府曾经试图阻止出版政府每年的税后收入报告（该报告表明富人越来越富，与政府宣称的相反），但是没有成功。政治也一直在影响人口普查所提出的问题以及对问题的解释（Alonso & Starr, 1986）。尤其是最近，社会安全管理部的专员因曾为一家信用贷款公司核查社会安全号码而辞职；对这些记录的保密是整个社会安全系统得到信任的基础。

研究过程的步骤

为了能在研究问题和档案数据之间寻找最优的匹配，建议采用如下步骤：

步骤1：确定问题

次级分析中的研究问题通常没有采用传统方法进行研究所提的问题那么具体。数据本身通常决定了能问什么问题以及如何来回答。你要做的第一件事情，除了确定你知道什么，还要确定对研究的主题已经了解了什么。

步骤2：寻找合适的数据库

要时刻记住不是所有从次级来源中收集到的信息都同样可信或有效。你一定要仔细评价数据，看看它们是否真正适合于你的需要。Stewart和Kamins（1993，pp.17-30）确定了评估数据时应该问的六个问题：

1. 研究的目的是什么？
2. 谁收集的信息？
3. 实际上收集到的信息是什么？
4. 信息是何时被收集的？

5. 信息是怎样被收集的?

6. 信息和其他来源的一致性如何?

记住, 当你为研究寻找数据时, 并不是只有这些问题。我们建议当使用次级来源时, 要保持一种“健康的怀疑主义”(p.31)。其他的限制可能也会存在。例如, 获得数据的过程可能会受到限制。有时原有的收集者(或者拥有者)可能要求在使用数据之前需要征求他们的同意。

步骤3: 研究计划的准备

次级分析研究计划的准备与其他类型的研究中所做的很相似。研究计划/报告的组织方法可能会随着读者(如大学教授或专业杂志)的不同而发生变化。对与研究问题相关的理论和经验性现有研究做一个全面的文献综述, 可以让你评价过去的工作, 并且发现需要探索的新的前沿领域。在研究计划的准备过程中, 其他人提供的信息能够辅助你选择研究方法(即发展测量方法、决定相关的样本、选择分析策略), 避免在努力中犯前面提到的错误。但正如任何一种研究方法一样, 需要权衡数据收集策略的优点与缺点。Elder等人(1993)提到的, “最有效的策略是充分利用数据的优势来形成个案”。

步骤4: 对档案数据最初的分析 and 重铸

一旦决定使用什么数据, 你就必须决定如何来使用这些数据。最初测量的时候, 数据通常不会完全适合于研究者当前的需要; 为了你的特定目的有时需要重新组织数据。这种重新组织, 有时叫做“重铸,” 是一个将一些变量重新分类并将那些含有不恰当数据的变量完全删除的过程(McCall & Appelbaum, 1991)。

步骤5: 分析

很明显, 数据分析的方法一定要适合问题、测量以及原有样本的特点。其实, 研究者往往不会提出那些无法用统计方法来回答的问题(McCall & Appelbaum, 1991, p. 916)。

总 结

虽然一个人可以列出研究的理想状态, 但真实生活的情境常会迫使社会科学家们止步于理想状态之前。因此, 典型的社会研究者面对的典型两难情境是, 要么不去研究他们想要研究的内容, 要么在与理想条件有差距的状态下去研究。研究者对方法的选择依赖于可以获得的资源。例如, 他们想从事一项调查, 但是发现资金并不允许, 于是就转向从事文献的研究或者使用一个早已经收集好的数据档案。一种方法往往更适于研究某

类问题而不是另一类问题。如果你想研究一个不发达国家中产生革命的原因，那么你一定需要对游击队员进行观察。相反，你可能使用已经存在的文件和记录，把革命与某些因素如贫穷联系起来。

档案数据为研究者从事重要的理论及与政策相关的、各科的问题提供了一个有效的、经济的且及时的方法。档案数据最大的科学价值也许在于它具有可被其他研究者们重复的可能。当研究在多种环境下可以被重复时，人们就会增加对关系有效性的信心。正如Traugott（1990）所述：

数据可以引发最初研究者远未预料到的新的分析。一项次级分析可能是因为一个研究者有一组不同的理论兴趣，或想使问题重新概念化，或由于档案中的数据积累到了一定程度。每一个原来为不同目的收集的数据，在合并或者一系列补充分析之后，可能会被重新分析产生新的价值。

如果研究者不得不组织自己的数据收集过程，他们在进行次级数据分析时就会得到更好的数据。他们可以把更多的努力用在分析上，而不是用在大规模数据收集项目的管理工作上。虽然在检查数据的原始收集者的概念和操作化定义时必须谨慎，但次级分析已经为我们增加了研究机会，扩展了我们对大量题目和问题的理解。这也正是科学更需要的——富于想像的、大胆的在一个不够完美的环境和不理想的条件下从事研究。研究方法的灵活运用让我们超凡脱俗，透过表面的现实更好地理解社会生活。

推 荐 阅 读

在这一节中，我们列出一些来自于心理学和社会学的研究文献用于进一步研究和深入阅读。Barrett（1994）、Davis和Smith（1991）分别是如何使用国家统计局和国家民意研究（NORC）一般社会调查的数据。这两种档案数据集都可被研究者得到，其优点是针对美国民众全体进行反复测量。Brooks-Gunn、Phelps以及Elder（1991）在《发展心理学》（*Developmental Psychology*）的一期特刊上，展示了心理学家如何对待次级（即档案的）数据，这一期特刊还包含其他一些相关文献（例如，Cherlin，1991；McCall & Appelbaum，1991）。一个经典的档案和纵向数据源是Terman和其同事对加州的高智商儿童的大规模研究，该研究开始于20世纪20年代，并且持续至今（Craven，1992）。其中最早的文献，Terman（1925）、Terman和Oden（1959）著作的第五版，可以提供有用的信息。最后，Young、Savelle和Phelps（1991）对社会科学中的纵向研究进行了探讨。

参 考 文 献

- Alonso, W., & Starr, P. (1986). *The politics of numbers*. New York: Russell Sage.
- Barrett, R. E., (1994). *Using the 1990 U.S. Census for Research*. Thousand Oaks, CA: Sage.
- Bastian, L., & DeBerry, M., Jr. (1994). *Criminal victimization, 1992*. Washington, DC: Bureau of Justice Statistics.
- Blau, J., & Blau, P. (1982). The cost of inequality: Metropolitan structure and violent crime. *American Sociological Review*, 47, 114-129.
- Brooks-Gunn, J., Phelps, E., & Elder, G. H., Jr. (1991). Studying lives through time: Secondary data analyses in developmental psychology. *Developmental Psychology*, 27, 899-910.
- Cherlin, A. (1991). On analyzing other people's data. *Developmental Psychology*, 27, 946-948.
- Craven, H. J., (1992). A scientific project locked in time: The Terman genetic studies of genius, 1920s-1950s. *American Psychologist*, 47, 183-189.
- Davis, J. A., & Smith, T. W. (1992). *The NORC General Social Survey: A user's guide*. Newbury Park, CA: Sage.
- Duncan, G. J. (1991). Made in heaven: Secondary data analysis and interdisciplinary collaborators. *Developmental Psychology*, 27, 949-961.
- Durkheim, E. (1966). *Suicide*. New York: Free Press. (Original work published 1897)
- Elder, G. H., Jr., Pavalko, E. K., & Clipp, E. C. (1993). *Working with archival data*. Newbury Park, CA: Sage.
- Gergen, K. (1973). Social psychology as history. *Journal of Personality and Social Psychology*, 26, 309-320.
- Inter-University Consortium for Political and Social Research. (1994). *Guide to resources and services, 1994-1995*. Ann Arbor, MI: Author.
- Kiecolt, K. J., & Nathan, L. E. (1985). *Secondary analysis of survey data*. Newbury Park, CA: Sage.
- McCall, R. B., & Appelbaum, M. I. (1991). Some issues of conducting secondary analysis. *Developmental Psychology*, 27, 911-917.
- Messner, S. (1983). Regional and racial effects on the urban homicide rate: The subculture of violence revisited. *American Journal of Sociology*, 88, 997-1007.
- Sampson, R. J. (1985). Structural sources of variation in race-age specific rates of offending across major U.S. cities. *Criminology*, 23, 647-673.
- Scott, J. (1990). *A matter of record: Documentary sources in social research*. Cambridge, UK: Polity.
- Stewart, D. W., & Kamins, M. A. (1993). *Secondary research: Information sources and methods*. Newbury Park, CA: Sage.
- Terman, L. M. (With others). (1925). *Genetic studies of genius, Vol. 1: Mental and physical traits of a thousand gifted children*. Stanford, CA: Stanford University Press.
- Terman, L. M., & Oden, M. H. (1959). *Genetic studies of genius, Vol. 5: The gifted group at mid-life: Thirty-five years of follow-up of the superior child*. Stanford, CA: Stanford University Press.
- Traugott, M. W. (1990). Using archival data for the secondary analysis of criminal justice issues. In D. L. MacKenzie, P. J. Baunach, & R. R. Roberg (Eds.), *Measuring crime: Large-scale, long-range efforts* (pp. 145-155). Albany: State University of New York Press.
- U.S. Bureau of the Census. (1964). *U.S. census of population: 1960, United States summary, characteristics of the population* (Vol. 1, Pt. 1). Washington, DC: Government Printing Office.
- Webb, E. T., Campbell, D. T., Schwartz, R. D., Sechrest, L., & Grove, J. B. (1981). *Nonreactive measures in the social sciences*. Boston: Houghton Mifflin.
- Young, C. H., Savelle, K. L., & Phelps, E. (1991). *Inventory of longitudinal studies in the social sciences*. Newbury Park, CA: Sage.

第五部分

研究报告的写作



第 21 章

APA 文体写作

——为何与如何

JAMES T. AUSTIN

ROBERT F. CALDERON

对科学界来说，通过发表文章来交流研究成果非常重要。如果没有交流，有价值的研究成果就可能无法影响到科学和社会。通过学术交流（Hills, 1983），研究成果在个人或单位之间得以共享，避免孤立隔绝。一般来说，审稿——出版的工作具备以下功能：传达信息、标定领域、设定质量标准、确认个人功绩、储存信息、丰富人们的精神生活以及保护公众利益。大学生和研究生常要准备实验报告或研究报告。他们在本科或研究生阶段的研究（学位论文）是学习如何交流研究成果的另一步骤。因此，写作是一个研究者职业生涯中一项重要的技能。

面向科学工作者的文体不同于学校所教的其他文体（Lannon, 1988; Sides, 1989; Woodford, 1967）。因此，心理学专业的学生不容易掌握这项技能。美国心理学会（APA）意识到这个问题，希望能规范心理学界的学术交流。所以，他们为研究者提供了一种特定的写作格式和文体，希望研究者在写研究报告、课题标书及其他形式的通讯时能遵从这种规范（APA, 1994）。这种规范也已执行一定时间了（Anderson & Valentine, 1944）。当你开始学习这种文体时，可能会觉得它陌生，但它是一种标准格式。按照这种格式，研究者能以逻辑严谨且简明扼要的方式表达理论观点和实验结果。

本章将介绍最近《出版手册》（第四版；APA, 1994）中规定的格式、文体的基本要求。因为计算机在出版中重要性日益增加，所以在讨论研究报告的各部分之后，我们一并介绍APA格式与文字处理技术（APA, 1994; Standera, 1987; Wright, 1987）。本章后的附录21.1是一模板，读者写文稿时可以它为参照，看是否符合出版手册。请注意，在此我们关注的是文体而非具体内容。准备草稿（Peterson, 第22章）和修订手稿（Nagata & Trierweiler, 第23章）的内容将在本书的其他章节介绍。另外，Becker（1986）、Runkel和Runkel（1984）以及Jones（1976）也详细介绍了科技界的写作方式。

尽管两者都要求作者自学，但写作符合APA规范并不等同于写得好。《出版手册》

为具体性 (specificity) 和敏感性 (sensitivity) 两个框架提供了一些建议 (APA, 1994, p. xxvi)。具体性是指详细描述研究细节, 以便让人能重复实验这一标准。敏感性是指不使用对某些社会群体有偏见的语言这一规范。APA特别推荐的可减少语言偏见的措施包括: 适度具体、注意标示的名称和过程是否过于敏感以及感谢研究参与者。其他的指导方针还包括: 精确选择用词、避免歧义、注意思想表达的顺序、从读者的角度考虑。以上这些都是有益的建议, 至于如何写作得更好, 本章后面所列的推荐阅读资料有更好的说明。改进写作文体与内容的方法之一就是实践。当你思考一个研究主题时, 写一些小方案 (2~3页), 将有助于你熟练掌握写作技巧的各个方面。《出版手册》中区分了三种类型的文稿: 研究报告、综述评论和理论研讨。你如果只是研究助理, 你不太可能去写后面两种论文, 因此本章将重点讨论前一种类型。

研究报告的组成成分

文稿一般由三部分组成: 前页 (front matter) 包括标题和摘要; 研究报告的正文 (body) 包括引言、方法、结果和讨论; 后页 (back matter) 包括一些材料证据: 参考文献、脚注、图表和附录。虽然并非每篇研究报告后页都要样样具备, 但研究者还是有必要了解一下, 并查阅《出版手册》以获得更详细的说明。此外, 正文四个主要部分中还可能包含一些细节 (例如, 方法部分可能包括参与者、器具/测量方法以及步骤)。下面我们按顺序说明每个部分。

引言

引言的作用是向读者介绍正在研究的问题和前人的相关研究。引言要回答读者的四个问题: (a) 你研究的要点是什么? (b) 与以往研究工作的关系如何? (c) 研究目的与实验设计之间到底有什么逻辑关系? (d) 本研究的理论局限性何在? 写引言最困难的部分是, 如何清楚地说明你研究的问题、以往的研究工作、你的研究设计之间的逻辑关系。你必须向读者解释清楚这些关系。

作者必须在引言即将结束时, 列出研究的假设或研究问题。假设是指对预期结果的特定预测; 研究问题更具有一般性、探索性。引言中的这个部分可能需要加独立的小标题, 而且在紧接每个假设陈述之前或之后, 都要有相应的解释。

总之, 引言为读者提供了一个判断研究目的是否与已有文献相关的方法。引言为正文中的方法、结果和讨论作铺垫。如果这部分写得不好, 读者可能会对你的研究失去兴趣或产生不良印象。因为编辑和审稿人往往是文稿的最初读者, 所以避免产生这种不良印象乃明智之举。

方法

研究报告的方法部分要详细说明你是如何收集数据的。它包含几个小节，研究的复杂程度不同，小节的具体数目也不同。一般而言，方法部分告诉读者研究的参与者（又称被试）、步骤、设计及所用的器具设备。一个合格的研究者在仅仅读完这部分之后，就应能评价此研究的质量，重复此研究。当然，你在阅读一篇文章的方法部分时，也可以这样“测试”一下自己。

参与者一节必须清楚解答三个问题：（a）谁参与了本研究？（b）有多少参与者？（c）这些参与者是如何筛选的（他们是否代表了特定的人群）？此外，须说明样本的主要人口统计学特征及分配到不同实验组或实验条件的人数。最后，还要交代给予参与者什么诱因（如现金、学分等）。

仪器装置一节应描述用于研究的主要设备或材料。如果是自制的装置，你还得给出更多的细节，以便读者能看懂。有时候，你甚至还要准备有关的附录或技术报告。问卷量表也应在这里提及，如果对原始量表或指导语作了改动，必须在此说明。有些人在这里说明心理测量学特性（例如信度估计、测量的标准差），有些人则在结果部分说明这些。

程序一节说明研究的实际操作步骤。它可能包括：指导语、参与者的分组形式及各种实验处理。描述清楚分组、平衡及其他控制手段。这一节必须告诉读者，你做了些什么、如何做的以及先后次序。再者，这部分必须足够详细，以便别人能重复该实验。在此，你还必须提及知情同意、事后解说及其他有关伦理上的要求。总而言之，方法部分就是详细说明参与者、仪器和测量方法以及刺激呈现和反应测量的程序步骤。

结果

结果部分总结收集到的数据及对它们的统计处理。开头要简要说明分析的主要类型，只有深奥复杂的统计检验才需要详细说明。然后你应详细报告这些分析，以便证明你将要得出的结论。而且，你应该只是陈述结果，不要讨论它们的含意或推论，因为这些都应放在讨论部分。你应报告所有有关的结果，包括与你的假设相反的结果（这点很重要）。

这部分中，你可能要用图或表格来表示结果，但要小心考虑这种表示方式是否值得，因为表和图并不容易排版（图比表更难排版）。常言道，“一图赛十文。”加入图表也是理之所在。如果你用了图表，无论是照片、图片还是示意图之类，均统称为“图”，各种表格均统称为“表”。比如，你可在文中说：“图1表示每个实验组对目标的接受水平。”如果文中含图或表，请把它们一起附在文后（参考文献之后）。表要有表头和页码，图则不必。但要在图前面的另一页上，标出插图说明（其中包含标题和页码）。《出版手册》中有关于图表的详细说明（表格一览表，3.74节，p.140-141；图一览表，3.86节，p.162-163），可供参阅。

数据的统计表达必须包含足量的信息，以便读者能确认你是否做了恰当的分析。因此，所有的推论性统计（如 t 检验、 F 检验、卡方检验）都必须包括有关信息，如均值、变异性（如方差、标准差）、检验所得的量值、自由度、证明虚无假设为真的概率、效应的方向性。例如：“正如预期，女性（ $M = 9.55$, $SD = 1.67$ ）报告喜欢学校的程度高于男性（ $M = 7.62$, $SD = 1.22$ ）， $t(22) = 2.62$, $p < 0.01$ ”“方差分析表明，没有接受训练的参与者其反应潜伏期显著不同于其他组， $F(1, 34) = 123.78$, $p < 0.001$ ”。然而如果你要报告详细的概率，就要在报告结果之前，先报告先验概率水平（即 α 水平）。例如：“所有统计分析的 α 水平均设为0.01”“当 α 水平为0.01时，性别效应显著， $t(22) = 2.62$, $p = 0.006$ 。”另外，如果描述性统计已经列于图或表中，那就不必在文中重复，除非它们支持/拒绝某个假设或研究问题。《出版手册》更详细地说明了如何充分表示统计所用方法及统计结果（1.10节，p.15-16；3.53-3.61节，p.111-119）。

越来越多的人将与假设检验有关的统计效力包含在此节之中。统计效力是指正确拒绝虚无假设的可能性，它依赖于所选的 α 水平、效应大小和样本大小（所有这些都应该在研究的概要中报告）。进一步说，在设计阶段，你就应该考虑到统计效力，这样才能做出适当的调整（如更大的 N 、增加试验次数），以保证研究有足够的统计效力。许多统计课程中都介绍了计算统计效力与效应大小的方法。

对新手来说，常用的分析方法包括：两组平均值的 t 检验，三组或以上的单因子方差分析，研究较大数据集合之间相互关系的一元或多元回归（Dickter & Roznowski，本书第16章）。高级复杂的技术包括多元分析，因为研究或研究生教学的普及而越来越普遍了（Levy & Steelman，本书第17章）。总之，结果部分包含了统计量总结、对先验假设和任何从你计划分析中产生的假设进行的显著性检验。

讨论

讨论部分是作者详细阐述结果的地方。讨论开始的方法之一就是首先表明数据是否支持了你原先的假设或研究问题。另外，你还要检查、解释或限定你的结果，提出一般的观点。讨论时你应考虑到以下的问题：（a）本研究有何贡献？（b）本研究对解决该领域中的问题有何帮助？（c）从本研究中能得出清晰的结论吗？（d）在该领域中还能做哪些进一步的研究？你需要适当注意一下结果的可靠性，也许还要应用Cook和Campbell的效度分类：内部效度、外部效度、统计结论以及结构类型，这些往往有助于自我评价。总之，讨论部分包含了结果概要、意义和自我批评。

文字处理与APA风格

鉴于应用计算机处理文本/文字现象越来越普遍，我们在此也说明一下如何建立默认

的格式。例如, Smith (1992) 为WordPerfect 5.1的用户提供了一些命令。不同的软件之间命令各不相同(如WP和MS-Word), 但几个主流软件所要考虑的问题, 无非都是行间距、页边距、页眉标题、对齐与断词以及拼写检查。首先, 稿件必须全用双倍行距, 参考文献和表格也不例外, 这是为了方便编辑和审稿人加评语。页边距至少要在2.5厘米以上, 只有页头例外。例外是因为2.5厘米是从顶边算到第一行文字, 而不是到页眉。因此页头最好留约1.5厘米。页眉通常在文字处理软件的“格式”中设置。页眉设置的两个要求是: 上/右对齐和页码显示(现置于与短标题同一行中)。文稿须用左对齐, 意即右边可以是参差不齐的。从初稿起就不要采用断词。最后, 我们强烈推荐使用计算机软件包提供的编写工具集。稿件在起草至送出去征求意见或审稿之前, 务必使用拼写检查, 这很重要。这并不是说你的拼写有问题, 而是有时候误拼乃心理运动(打字)所致。辞典(thesaurus)将有助于避免作者滥用某些词。最后, 语法检查还能发现诸如滥用被动语态之类的常见错误。另外, 工具集还能呈现有用的统计信息(如公式、字符/字/段落数)。我们建议: 从开始使用文字处理软件的那天起, 你就学习APA格式, 并不断练习, 使书写格式变得自动化。

有几个商业软件有专业编写功能, 它们中有一些还有管理书目功能, 以APA格式产生参考文献列表(例如, End Note Plus, Niles & Associates, 1993; WPCitation, Oberon Resources, Inc., 1989), 而另一些软件则给出整个文档的框架(例如, Manuscript Manager, Pergamon Press, Inc., 1988)。

总 结

总之, 经过努力, APA文体与格式还是比较容易学的。在习惯书写这种格式之后, 你会更多地为老师和研究伙伴所青睐。而且, 使用这种文体也显示了你的专业素养。本章围绕研究报告文稿及附录21.1中的样例, 简要讨论了第四版《出版手册》中的一些要点。当然, 你得仔细阅读《出版手册》及相关材料, 除此别无他途。任何人只要反复练习、应用就能写好文体。所以, 当为班级或实验室写稿时, 要习惯于反反复复、周而复始地写。尤其是为班级修改论文、准备小方案时, 你需要不断地练习、练习、再练习。

推 荐 阅 读

本章最后一节推荐一些阅读材料。这些读物包括《出版手册》本身, 它是最重要的, 也是最权威的(Knapp, Storandt, & Jackson, 1995)。Hummel和Kaeck(1995)根据《出版手册》写了短文介绍APA格式的精要。但是, Chapell(1995)批评了Hummel和Kaeck(1995)所描述的格式, 因为后者的依据是第三版手册, 所以格式有点出入。

更一般的关于文体的三个经典读物是：《文体要素》（*The Elements of Style* <3rd ed>; Strunk & White, 1979）、《芝加哥文体手册》（*The Chicago Manual of Style* <14th ed>; 芝加哥大学出版社, 1993）和《文体手册》（*The Manual of Style*; 政府印刷办公室, 1986）。其他为社会科学写的书本有：Becher（1986）、Jones（1976）、Runkel和Runkel（1984）以及Sternberg（1993）。Rubens（1991）写了一本文体手册，特别适用于从事科技写作的作者。最后，Sternberg（1992）的文章提出一个大纲，论述如何为心理学杂志写稿，大纲主要包括以下几个要点：（a）你要说什么；（b）如何说；（c）如何处理你所说的内容；（d）如何处理他人（如同事、评论者）所说的内容。

参 考 文 献

- American Psychological Association. (1994). *Publication manual of the American Psychological Association* (4th ed.). Washington, DC: Author.
- Anderson, J. E., & Valentine, W. L. (1944). The preparation of articles for publication in the journals of the American Psychological Association. *Psychological Bulletin*, 41, 345-376.
- Becker, H. S. (1986). *Writing for social scientists: How to start and finish your thesis, book, or article*. Chicago: University of Chicago Press.
- Chapell, M. S. (1995). Correction to APA style. *American Psychological Society Observer*, 8(6), 33.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin.
- Hills, P. J. (1983). The scholarly communication process. *Annual Review of Information Science and Technology*, 18, 99-125.
- Hummel, J. H., & Kaeck, D. J. (1995). How to use the '94 APA style guide. *American Psychological Society Observer*, 8(4), 16-22.
- Jones, W. P. (1976). *Writing scientific papers and reports* (7th ed.). Dubuque, IA: William C. Brown.
- Knapp, S., Storandt, M., & Jackson, D. (1995). Fourth edition of the *Publication Manual*. *American Psychologist*, 50, 581-583.
- Lannon, J. M. (1988). *Technical writing* (4th ed.). Glenview, IL: Scott, Foresman.
- Niles & Associates, Inc. (1993). *End Note Plus* [Computer program]. Berkeley, CA: Author.
- Oberon Resources, Inc. (1989). *WPCitation* [Computer program]. Columbus, OH: Author.
- Pergamon Press, Inc. (1988). *Manuscript manager* [Computer program]. Elmsford, NY: Author.
- Rubens, P. (Ed.). (1991). *Science and technical writing: A manual of style*. New York: Holt.
- Runkel, P. J., & Runkel, M. (1984). *A guide to usage for writers and students in the social sciences*. Totowa, NJ: Rowan & Allanheld.
- Sides, C. H. (Ed.). (1989). *Technical and business communication: Bibliographic essays for teachers and corporate trainers*. Urbana, IL: National Council of Teachers of English, Society for Technical Communication.
- Smith, R. (1992). Formatting APA pages in WordPerfect 5.1. *Teaching of Psychology*, 19, 190-191.
- Standera, O. (1987). *The electronic era of publishing*. New York: Elsevier.
- Sternberg, R. J. (1992). How to win acceptances by psychology journals: 21 tips for better writing. *American Psychological Society Observer*, 5(5), 12.
- Sternberg, R. J. (1993). *The psychologist's companion: A guide to scientific writing for students and researchers* (3rd ed.). New York: Cambridge University Press.
- Strunk, W., & White, E. B. (1979). *The elements of style* (3rd ed.). New York: Macmillan.
- University of Chicago Press. (1993). *The Chicago manual of style* (14th ed.). Chicago: Author.
- Woodford, F. P. (1967). Sounder thinking through clearer writing. *Science*, 156, 743-745.
- Wright, P. (1987). Reading and writing for electronic journals. In B. K. Britton & S. M. Glynn (Eds.), *Executive control processes in reading* (pp. 107-144). Hillsdale, NJ: Lawrence Erlbaum.

附录21.1 APA格式写作示例

以下几页是APA文体写作的一个两层次模板，是从Evan Mendes博士（George Mason 大学）和Peter Villanova博士（Appalachian 州立大学）的早期版本演变而来的。本附录有两个层次：第一是它的物理特性，即现行APA格式（1994）的真实手稿格式，本附录的文档是以实验研究报告的格式组织的；第二是认知特性，与具体应用它的物理特性有关。例如，模板的摘要部分既描述了摘要的一般原理，也介绍了写摘要的详细步骤。因此，本附录中的文档既用了APA格式，同时其内容也是有关正确使用格式的。关于本附录，作者声明两点：首先，一篇研究报告中所有的特性不可能都包含在内，因而《出版手册》仍是最后也是最好的裁判；其次，页码从1开始。

页眉标题：你选择的短标题（均大写）

1

你的手稿标题

（大小写混合，居中）

你的姓名

所在单位

（James T. Austin编写）

通讯

姓名

地址

电话和电子信箱地址（不一定必要，但很有用）



摘 要

本范文用两种方式教你使用美国心理学会（APA）文体。文本解释一篇典型研究报告的要素和组织方式，同时本身也符合正确的APA格式。你正在阅读的这部分，也就是摘要，在紧接标题页后另起一页（在手稿中加下画线的作用是：告诉编辑和排版工画线部分用斜体，而不是在文字处理软件中直接用斜体）。该页必须在页面的右上角写页眉标题，后接页码（“第2页”），摘要一词居中放置。摘要以块状方式排版，另起一段，段首无缩进。摘要包含文章的概要。它应该给读者提供信息，以便他们知道这篇文章是否与他们的兴趣有关。在许多情况下，读者在决定是否读完全文时主要是依据摘要。因此，它应该具备关于问题、方法、结果以及推论和结论的信息。特别注意的是，摘要应在100~150字之间，不超过960个字母（这个特殊的摘要205字，有点长了）。总之，摘要要言简意赅。



你的手稿标题

（大小写混合并居中）

在标题页和摘要之后另起一页写引言部分。注意，页眉标题和完整标题要出现在引言第一页的页顶，但没有副标题。另外，全文排版，不右对齐，也不要断词。读者可以参考美国心理学会出版手册（第四版；APA，1994），以获得其他有关手稿写作的信息。其他关于社会科学写作的注意事项可以在Sternberg（1993）和Becker（1986）的书中找到。Strunk和White（1984）的《文体要素》是写作一般文体的有用参考资料。这些资源列于本章的参考文献部分。

因为引言的功能显而易见，所以小标题不是必需的。具体说来，引言告诉读者正在研究的问题以及这一问题与以往研究工作的关系。所以在写引言时，你心中要问四个问题：（a）本研究的要点是什么？（b）本研究与该领域中以往工作的关系如何？（c）研究的问题与研究设计之间有何理论上或逻辑上的关联？（d）本研究有什么理论局限性？写引言最困难的是如何清楚地表达你研究的问题、以往的工作以及你研究设计之间的逻辑关系。写稿时要把这些问题说清楚。

引言要讨论文献，但不需要过多的综述。如果你已经做了很多的综述，可以考虑把它写成综述/理论研讨文章、定量综述（元分析）、有评注的书目或技术报告。当写一篇研究报告时，尝试只引用那些与当前研究问题有关的文献。当为你的文献综述选择文章时，既要包含最近的文献，也要包含经典文献，前者确保你不是老调重弹，后者显示你掌握了这个现象的来龙去脉。如果所研究的领域里已有（通常情况下）元分析（定量综述）文章，请考虑引用它们。不要完整描述你所引用的研究，你常常只需简要描述一下变量、程序、分析及发现，还有结论。你要避免离题或笼统引用，除非它们支持你的讨论。

在论文的正文部分，列出作者的姓氏及发表年份来引用文章或书籍。这种形式的引用称为科学引用法，以区别于文学作品中的引用形式，后者依从的是现代语言学会（MLA）制定的规范。例如，你可能会说：“Skinner（1938）最早提出了操作性条件反射与经典条件反射之间的区别，但现在有人对此提出了挑战（Rescorla & Wagner，1969）。”注意，当引用是在括号内时要用“&”号代替“和”字。但在括号外边还是用“和”字。引用同一作者同一年的不同文章或书籍，就在年份后面加上后缀a、b、c等等。这些后缀同样用在参考文献部分。例如，你可能会说：“最近Smith（1965，1975，1978a，1978b）的研究支持了期望解释。”如果在一个地方引用不同作者的作品，那么就引在括号内按作者姓氏的字母顺序排列，以分号隔开：“最近的研究质疑了行为主义者的地位（Smith，1983；Smith & Jones，1984；Toffler，1979）。”

尽量不要用脚注和引文。你引用的文献必须来自原作，而且是用你自己的语言转述。剽窃是极端不道德的，多数学校和单位对此行为均有重罚。然而，有时你别无选择，只能依靠二手资料，即别人的引述。当你必须引用二手资料时，应把原始来源放在文中，如“Heider（1958）报道Asch观察到……”在参考文献部分，列出Heider，而不是Asch，尽管我们感兴趣的是Asch的文章。

写文章时要尽量明白、清晰，不要用华丽的辞藻。例如，“美国杰出的临床心理学家，George Kelly博士，因率先把个人描绘为幼稚科学家而著称。”最好还是表述成，“Kelly（1955）率先把个人描绘为幼稚的科学家。”写作的三个简单原则是：了解读者，了解交流的目的，简单、简短、准确。提纲可以帮助你理清思路，你可在写完之后（反拟提纲）或开始写作之前拟订。

假设

作者应该在引言部分的最后列出假设或研究的问题。假设是指对结果的预测；研究问题是概括性的，表明这个研究是探索性的研究。引言这部分也可能用到三级小标题。

总之，引言告诉读者，你为什么要在现有的研究框架中做这个研究。它为整篇文章包括方法、结果、讨论及参考文献部分作了铺垫。

方 法

方法部分以方法一词开头，大小写混合并居中。此列为一级标题，紧接在引言及假设或研究问题之后。方法部分通常再分几个小节，具体分几小节根据研究的复杂程度而定。它的目的是告诉读者关于参与者、所用程序、所用设备的信息，以便同行能评价并重复该研究。

参与者

方法部分中的各小节（如参与者）有大写的标题，有下画线，并且左对齐。这些在APA文体指南中称为三级标题。参与者小节应该清楚回答三个问题：（a）谁参加了本研究？（b）有多少参与者参加？（c）参与者如何筛选（他们是否代表了某个特定群体）？另外，你还需列出主要的人口统计学特征以及每个实验组或实验条件的参加人数。要报告给了参与者什么诱因，如现金或其他。最后，被试一词有时会让人觉得有贬损的味道，所以可用参与者一词取代。

设备/测量方法

在此小节，简要介绍研究中所用的主要设备或材料及其功能。如果是自制的装置，你得给出更多的细节，以便读者能看懂。有时候，你也许还要准备有关的附录或技术报告，问卷量表也应在这里提及。另外，对原始量表或指导语所作的任何改动，都必须在此说明。

程序

该小节表述了实际研究中的每步工作，包括给参与者的原始指导语，实验组的编排，以及特殊的实验处理（如果有的话）。你需要描述清楚分组、平衡及其他控制手段，并按照APA原则报告对参与者的事后解说。一般说来，这部分告诉读者你做了些什么、如何做以及按什么顺序做，要详细到让他人能够重复实验。



结 果

结果部分紧接在方法部分之后。结果一词用一级标题，居中，大小写混合。该部分的目的是总结数据，进行统计处理。首先，你要简要说明结果或发现的新突破之处。其次，你要详细报告数据分析，以便能在后面证明你的结论。你只需说明结果，暂不要讨论其意义或做出推断，这些均留到讨论部分。你应报告所有有关的结果，甚至那些可能与你的假设相反的结果（这点很重要）。

你可能会选择使用图或表来表达数据或结果，一方面，请仔细考虑这种形式是否有价值，因为表和图是不易排版的；但另一方面，常言道“一图赛十文”，加入图表也是理之所在。如果你用了图表，无论是照片、图片还是示意图之类，均统称为“图”，各种表格均统称为“表”。例如，你可以说：“图1表示每个实验组的目标接受水平。”如果有图或表，你要把它们都一起放在手稿的结尾处。对于图，你需在紧靠图的前面单独一页上写上标题说明（表和图的示例在这个指南的后面），但要在文中指出表或图的位置，如下：

在此插入表（或图）1

当你报告统计显著性检验的结果时，必须交待均值、变异性（例如方差、标准差）、检验的量值、自由度、虚无假设为真的概率、效应的方向等。这里有几例：“正如预测，女性（ $M=9.55$ ， $SD=1.67$ ）相比男性更加喜欢学校（ $M=7.62$ ， $SD=1.22$ ）， $t(22)=2.62$ ， $p<0.01$ 。”还有“方差分析发现了没有参加训练的参与者在反应潜伏期的显著性差异， $F(1, 34)=123.78$ ， $p<0.001$ ”。报告统计显著性及效应大小的惯例在统计学和研究方法课程中有所涉及。表格常用来表现复杂分析的结果，诸如因素分析、多元回归或因素方差分析。

在读本科时，最常用的统计分析方法包括两组非独立变量均值的 t 检验、比较三组或更多组在某个因素上均值之间差异的单向方差分析、两组变量的简单相关分析和回归分析、多组变量间相互关系的多元回归。更多的分析技术会在高级心理学研究中有所涉及。

讨 论

讨论部分紧接在结果部分后面。标题居中，大小写混合。一种开题就是陈述结果对你原先的假设或研究问题的支持。另外，你可以检验、解释、修饰、限定你的检验结果，同时引申出其普遍意义。讨论时要考虑到以下问题：（a）此项研究的贡献是什么？（b）我的研究有助于解决该领域中的问题吗？（c）我能从此研究中得出清晰的结论吗？（d）这个特定领域还能做什么进一步的研究？

总之，研究报告遵循的是信息的逻辑顺序，从提出问题、构思问题，到研究问题，再到分析数据、做出发现、解释发现以及对以后研究提出建议。最后，请记住，提高写作水平的最佳办法是练习、练习、再练习！

（另起一页开始参考文献部分）

参 考 文 献

心理学文章把参考文献放在文后，这样做不会干扰手稿的阅读。一般情况下，你应很少在文中作注（它们也放在参考文献部分，文献之前）。这部分紧接讨论部分之后另起一页。列参考文献的目的是让读者获得文中的参考资源。参考文献（标题）应大小写混合，居中对齐。以下是手稿参考文献通常所要求的格式示例。有关参考文献的信息与示例，请参看APA出版手册（1994，附录3-A，pp. 189-222），也可找到有关引用合法材料的说明（附录3-B，pp. 223-234）。

Adams, J. E., & Aronfreed, J. L. (1971). Group dynamics in the prison community. Journal of Experimental Social Psychology, 42, 17-21.（文章示例）

American Psychological Association. (1994). Publication manual of the American Psychological Association (4th ed.). Washington, DC: Author.（APA文体指南——“圣经”）

Riesen, A. H. (1966). Sensory deprivation. In E. Stellar & J. M. Sprague (Eds.), Progress in physiological psychology (Vol. 1, pp. xxx-xxx). New York: Academic Press.（书籍章节示例）

Stogdill, R. M. (1974). Handbook of leadership. New York: Free Press.（书籍示例）



附 录 A

自选的、一目了然的标题

（居中，大小写混合）

附录是评审者或读者感兴趣的额外资料。这种材料不方便放在正文中，但对阅读或评论手稿来说却很重要。这种资料包括：作者为研究项目特别制作的问卷、计算机程序清单、刺激材料列表（如言语学习实验），或者是复杂的数学证明。书写请用双倍行距，在参考文献之后另起一页，用不同字母区分每个附录。如果只有一个附录，就不需要字母。

作 者 注

在典型的研究报告中，作者注包括三个部分：第一是作者的单位部门；第二是致谢，感谢同事、其他研究机构或研究助手对你的帮助；最后一部分指明如何与作者取得联系，包括标准地址和电子信箱。作者注不编号。



表1

先天营养不良的大鼠与控制组走迷宫平均成绩

表注在此

（包括统计显著性，后接其他解释）

图1

（为图编一个简短而一目了然的标题）

从出生至200天，先天营养不良组的平均体重（X……X）与控制组的平均体重（O……O）。

（图应在此说明之后的另一页上）



第 22 章

打 草 稿

Christopher Peterson

给撰稿者提建议的所有人都强调：最终的作品莫不是几易其稿而成。尽管这个建议是老生常谈，但许多本科生和研究生还是经常不打草稿。本章的目的不是要鼓吹草稿的价值（我认为它是理所当然的），而是介绍写草稿的一些具体步骤。

奇妙的思维

正如许多弄巧成拙的行为一样，写作出现困难部分是由于坚持并未得到检验的信念（Beck, 1976）。习惯化的思维常常支配着人们，一些顽固而又过分简单的信念莫名其妙地控制了头脑，导致人们在情绪上和行为上出现问题。奇妙的思维维护着这些想法，使得人们无从对其进行真实性检验。对于写作来说，奇妙的想法只与开始动笔和最后完稿有关，因为它不关心中间的过程。人们谈及写作时，常常用到艺术、灵感、天才等字眼。实际上，写作只是一种平凡的活动，需花费一定的时间。不断地重复尝试乃是写作的本质。

让我说得具体一点。一个满怀热情的撰稿者也许会认为，写作无需努力，而当事实证明并非如此时，他就可能会认为自己不懂写作。在校学习生涯早期的社会化对学生来说似乎并不太好，因为它使我们或多或少地觉得，“好”学生不费劲；“勤”只是用于补拙；快等于聪明；事情要么得来不费吹灰之力，要么徒劳无功。这种刻板印象与真正的写作丝毫沾不上边。

一个撰稿者可能会认为，遣词造句有十全十美的方法。这样，如果语不惊人，或不能一挥而就，那么结论就是此人还不足以胜任写作。确实，一个人总无法动笔，除非他确切知道自己想说什么。如何做到胸有成竹呢？阅读相关主题的所有文章和书籍是一种“策略”。这时，学生就会一头扎进图书馆，动笔的事无限期拖延（见本书第2章）。

我经常问那些准备写作的人，当他们写作时，是谁钳制着他们的思想。有的回答说

是严厉的父母；有的说是七年级的老师，说到写作时总是满口的分词、过去完成式、图解的句子；有的回答是位受人尊敬的人物，这位人物在不经意间把写作过程弄得很神秘；而更普遍的回答是，有一个不认识的、却又无所不知的“他们”强行武断地裁定一个人的智能。

多数学生最终还是写了，但因为写作似乎与自尊和能力有千丝万缕的关系，所以他们往往对自己的作品感到不满意。其实，真正妨碍学生写作的，是他们自以为的必须像宗教仪式般地执行某些步骤的念头（Becker, 1986）。我并非谴责写作的例行步骤，因为它们显然是必需的；我只是反对相信写作有一个奇妙步骤的想法，因为它们妨碍写作者动笔。

相信自己能写得出来的信念，也许就是决定能否写出有用草稿的关键因素。这个信念受几个因素影响，包括以往成功完成草稿的历史、他人成功的例子、有人苦口婆心地告诫你草稿定能写成（Bandura, 1986）。本章就是上述最后一个因素的实例，外界的说服通过自我效能感对本章的写作产生了作用。诸位读者应该知道，现在你看的本章是从起草以来的第六稿。

主人翁精神

学生们经常不明白自己写作的目的。当然，在某种意义上，我们是为某些读者而写作，而且通常也是由于外界的要求：教师布置写作文；研究生录取委员会要求有个人陈述；研究生导师要求写研究方案；学术职业要求发表文章或专著。但写作的最直接原因应在于作者心里，准备写作的人应该首先了解这一点（见本书第1章）。

对正在做的事，我总是鼓励学生以主人翁自居。写作的主人翁精神最重要的就是担起责任。“写作”是个动词，要有人来执行。如果是你的文章，那个人就是你。这种想法随着作品写作过程不断延伸，变得越来越有必要。

我的博士生在说到学位论文时，几乎总在论文前加个“the”。我的天啊！这就好像在谈论天气或政府一样，总觉得它们是非己所有、不可控制的。把某人的写作归入这类，那就等于把个人在写作过程中的作用完全阉割掉。

我坚持要求我的博士生在谈及学位论文时不许用定冠词，而用人称代词。我甚至叫他们压根就不要用“学位论文”一词，因为它听起来太过堂而皇之。从本质上讲，他们写的并非学位论文，只是字词章句。当一个学生以适当的方式写下字词章句时，我们才称之为学位论文。当我的学生开始谈及撰写他们的研究报告时，我知道这个过程开始了。当他们开始谈及写下的假设、方法和结果时，我知道这个过程正在顺利进行。

一种神秘的看法是，有些学者只写好文章。由此推论，有些极优秀的学者只写了为数不多但极具影响力的文章。但事实证明并非如此。写出引用率很高的文章的人也是写

了很多文章的人（Merton, 1968）。这些文章有些极具影响力，有些则没有。想想弗洛伊德、勒温和斯金纳这些心理学界的泰斗吧。他们有些文章也是笨拙蹩脚、语无伦次、不合逻辑的。但我们只推崇他们的好作品。我敢肯定，他们没有一个一开始就想写成经典的著作，他们只是在不断地写。

我的论点是：撰稿者不要同时扮演评论家或历史学家。是的，作者应该努力写好文章，但这只要投入时间和精力。不要削足适履以追求不切实际的深邃或轰动。

下笔如涌泉

显然，草略是草稿的特点：初始、未完，甚至愚蠢和自相矛盾。许多学生不愿意写草稿，因为“草”字具有负面的涵义。结果，作者往往在开始动笔时就想写出成品。每句话都必须合乎语法，每个例子都必须切题，每个转折都必须天衣无缝。这些目标让人望而生畏，如想一下笔就成完美文章，那你很快就会停止不前，连第二句都写不下去了。

较现实的方法是把写稿和修改分开。撰稿者可以采用涌泉法，在写第一稿时让字句随思绪翻滚而出。涌泉法的唯一规则是一直往下写，不要回头修改。当然你还是需要一些正常的判断力，因为你要写的是某一主题，所以要确保不离题。除此之外，一切应任其自然。

换言之，撰稿者只需对要写的内容有个大概想法就可开始动笔了。我认为，人们过分强调了诸如做提纲计划这些观念。当然，大致的提纲很有用，我写作时也用它。但如果在真正动笔前就一定要写出详细提纲，那么拟提纲的过程中，就容易出现完美主义倾向，这样撰稿者要完成作品的话，就会多了一重障碍。

我们所写的心理学文章多数有公认的格式。例如，一篇研究报告包括引言、方法、结果和讨论几个部分。每个部分都有固定的组织方式。一旦写作者头脑中有了这种组织方式，那么在他开始动笔前还需要进一步的大纲吗？

写文献综述方面，我经常劝我的学生读写交替。在写之前，他们一定要熟悉主题，但不必读遍所有文献。我强烈主张不必阅读所有文献，读到足以动笔的程度即可。在写作过程中，他们写到需要更多资料的地方时，可用星号作标记或在括号内作注。这时，作者一定要继续文思如泉涌般地写下去，随后再来搜索所需的详细资料。

我经常为要写的主题作笔记，这是准备工作的一部分，而且我总是要阅读很多文章和书籍。但我在打草稿时，绝不会查阅这些资料，除非我对要表达的意思有疑问。当我在修改、润色时，才把它们派上用场。

打草稿时不必固守顺序。人们在阅读最终版的文章或书籍时习惯从头到尾，但写作时不一定按此顺序。我常常从最熟悉最轻松的部分开始写，拿起笔就写，不管这页内容

最终将放在文中何处。这是开始起草的好办法。

例如，撰写一篇研究论文时，先从方法部分写往往是最容易的。我通常都这样做，因为这部分最按部就班。然后我跳到引言的结尾，提出我的假设，接着是起草结果部分。我通常在完成数据分析之前就写完这些，这样效率比较高，因为当我试图描述一个研究表明了什么时，我才清楚我需要哪些数据。在写结果之前，我们通常不是马上就清楚需要做哪些分析。实际上，不需做过多的数据分析，这很浪费时间、计算机资源和资金。

你可用许多办法来协助实施涌泉法。读本科时，我总是撇开笔，直接在打字机上打初稿。虽然我打字是“二指禅”，但打字还是比手写快，而且我有意尽量不删改字句。随着文字处理软件的出现，打草稿变得更容易了。当然，文字处理软件也常常诱使人玩弄拼写与措辞，而不是真正写作。

如果作者抵制不住修改的诱惑，一个极端的应对措施就是关掉显示器，往电脑里狂打文字。因为你看不见，自然不会去修改。当然，你得确认没选文字处理软件的改写选项。另一个极端的做法是口述录音，然后翻录成文字。这比较花时间，但如果你写初稿真得很有困难，这不失为权宜之计。

当学生们问及文章该写多长时，有些老师表现得很不耐烦。我觉得这是误导。在现实中，写作需符合页面限制，知道作品该写多长才合乎情理。当我开始起草时，尽管不太清楚要写些什么确切内容，但对于要写多长的文章却不会有什么含糊。对其他计划写作的部分也是一样。文字处理软件可监视已写了多少字，因为它们一般都有字数统计或页数统计。

搁 置

草稿完成后需要修改，但在修改之前很重要的一点是：先别动它。把它搁置一两天以上，以便作者可以用全新的眼光来看待作品。我们往往需要一定时间去酝酿其思想及表达方式（Poincare, 1913）。我的写作生涯中常有这样的体验：当我回头看一篇草稿时，搞不懂当初我为什么这样说而没有那样说。

搁置草稿有另一好处：当作者回头看时，会发现它看起来要比原先完整得多。我写引言时总是写不太好，尤其是开头的文献综述往往过于冗长（见本书第24章）。我努力要描述问题的发展，但我常感到过犹不及。我想了一个解决办法，就是先把引言写得很简短，等待修改时再来扩展，而事实多半是，我最终决定这么简短的引言就已足够。

隔多长时间写下一稿得有时间限制决定，这里我们要牢记一句格言：尽快开始执行计划，要意识到拖延的危害。讨论拖延（Silver & Sabini, 1981）已超出本文的范围，但预先计划给其他人看稿子，是尽快起草以备搁置的好办法。作者必须确定一个最后期

限，给他们看稿，并让自己感觉到，这个期限和真正的交稿期限应是一样重要的。

修 改

假设时间已过，作者回到草稿上来准备修改，那么这时就应有个如何写下一稿的计划。作者也许需要继续用涌泉法来写作，也许需填补空白、补充资料，也许需检查语法、标点和拼写，一篇草稿太草略，所以他同时要做许多事情。如果作者想一蹴而就，把草稿变成成品，他就可能会陷入困境。

文章写得越来越多时，人们通常会意识到自己写作的缺点，而修改时就须特别注意这些缺点。例如，我在写初稿时常写被动句，当我修改时就找出这些句子，把它们变成主动语态。我还写很多以thus和indeed开头的句子，这些在我看来都只是浪费空间。当我修改时，我查找这些单词，删掉其中的大多数。

还有一些缺点本质上属于文章结构方面，也要用以上的方法来对待。作者在写草稿时，可能一贯写得太多或太少；可能会忽略举例或转折；可能没有说清楚结论的重要性。作者应该列出这些通病，在修改草稿时要特别注意修正这些毛病。多数情况下，这个过程会变得自动、熟练，融入打初稿的过程中。

尽管我反对为初稿拟提纲，但拟提纲确实有益于修改过的草稿（见本书第23章）。反拟提纲是个好策略，它为已写的内容拟提纲。反拟提纲可鉴别出放错位置的段落或小节，可修正不均衡的分布。

接下来谈谈是保存前稿备份还是删掉它。有些作者使用字处理软件，在同一张磁盘里保存了同篇论文的十几个版本，然后搞不清楚到底是哪个。学生也常常错拿论文稿给我，有时还会误把前稿覆盖后稿，有时会把论文的几个不同版本混杂一起。

我的做法是不保留已修改的草稿备份，因为我认为我绝不会重用它们。如果我觉得一段或一页内容可能日后会有用但不适于现稿，我会在手稿的最后另起一节保存它们。我认为，既然毛病已经改掉，就没理由再保留充满毛病的稿子。如果你不想听我的，那至少要保证前后稿件要能识别清楚。或许，用稿一稿二这类号码或完成日期来作标记。

听 从 建 议

写作并不像世俗认为的那样，是闭门造车。在撰稿时，作者除了要广泛阅读，还要让其他人参与你的工作，提出不同意见，以迎合不同的读者。许多人不肯这样做，因为草稿的拙劣让人尴尬，就像丑媳妇怕见公婆。但不管怎样，听从别人的建议是写作的一个关键步骤。

准备写作的人要战胜自己，不要怕别人提意见。应该意识到：在可更改的情况下听

到批评，总比在无可挽回时听到批评要强得多。多数老师更愿意教导而不是给你打分，他们通常乐于在草稿上写评语。研究生导师也是如此。我把投稿或标书的初期版本交给杂志编委或出资单位，有时居然也能顺利通过。同学或同事最合适提建议。我总是固定让一个同事给我的文稿提建议，彼此互相帮助。再强调一遍，我认为每个作者都应为其草稿诚恳地征求意见。

并非每个看草稿的人都能对你有帮助；他们的建议有的来得太迟，有的太挑剔，有的太草率。批评者可能指出某个句子不好或不清楚，但不会解释为什么。作者应该四处寻觅，寻找好的批评者，而且一定要把你需要什么样的建议清楚地告诉他。要警惕好好先生，因为无条件的赞扬只会让作者自我感觉良好，对改进写作无益。

一旦有人愿意批阅你的稿子，你应做些什么来让建议者心情愉快呢？首先，自己能做的事不要依赖别人。例如，把别人当成拼写检查软件是种侮辱和浪费。如果人家花了许多时间为你更正拼写错误，那代价就是少了许多实质性的建议。

有时学生给我一篇草稿，让我提建议，但交过来的稿件印得密密麻麻，顶天立地，没有边缝。这样不仅读起来很费劲，也没地方写评语。其实，单从稿子的外观就能看出撰稿者对待建议的态度。同样，别将一张存有草稿的磁盘交给建议者，期望他先打印出来再写评注，除非你们俩事先商量好如何处理。再者，如果你们使用的是不同的操作系统或字处理软件，那你就会浪费建议者更多时间。

虽然草稿不必要完整，但也不宜太粗略。拿给人批阅的稿子至少要有些内容。有时学生拿出一个干巴巴的骨架提纲（开头、中间和结尾）让我过目，而我却通常摸不着头脑，不知该说些什么。如果作者想要别人为文章的结果部分提出实质性建议，就应把基本的表格和图示附在里面。我很少看到有人把摘要也附在草稿中给我评注，但其实摘要对草稿和终稿的读者来说一样有用，甚至更有用。

如前所说，告诉别人你最需要什么样的建议，越具体越好。还要约定交还评注的最后期限。你也许无法控制建议者何时践约，但也不至于说好周一拿，却到周五还没拿到手。

当你收到返还的建议时，要取其要旨而改之，别只看到文稿上批注的地方而忽略其余。记得有个人，我在他的学位论文里指出20个相同的单词让他改写别的，他老老实实在我指出的那20个地方作了改动，其他的重复单词他就不管了。所以，我的意思是，要仔细参详别人的建议，如果有帮助，应从中吸取教训，推广应用。

当然你应感谢给出建议的人，替他看草稿也许就是最好的致谢途径。如果建议被采纳，那你就坚持做下去。所谓坚持到底，是指你能认真对待。这并不意味着凡事都把别人的建议当令箭，但也不是说为不采纳建议找借口（见本书第25章）。至少，如果某人建议说稿子应该这样改，作者通常应或多或少改一下。因为如果没有毛病，就不会引起别人批评。

我请求建议者尽量在文中修改：重组、选词、标点，诸如此类。我基本上都是综合了这些修改，但如没能这样做，我会仔细考虑问题的实质，寻找修正的方法。如果别人不喜欢我用的词，而我又不喜欢他提议的词，那我就会找另一个。因为接受建议不是参加辩论，作者不采纳诚恳的建议并不等于战胜论敌。

假如你很幸运，可以有几个人为你提建议，那么时间允许的话，让他们逐个提议是利用这些有利资源的最佳办法。你同时收到的不同建议往往互相矛盾，让人无所适从。正如写稿和修改，铺开收集的建议和修改的过程是最好的方法。

总 结

总之，我觉得打草稿是一个需要时间的过程。作者需要区分打草稿的几个步骤并分别执行。首先，作者对写作要有主人翁精神，要对写作的后果负责；其次，作者需把其大致思想写于纸上；第三，如果可能，把草稿搁置一段时间；第四，草稿必须重写，也许还要大改几遍；第五，作者需寻求、听从别人的建议。

打草稿需提前进行，留一段时间作最后润色。无益的设想和习惯都会妨碍打草稿，作者应随时警惕，注意克服。但就算已做足了这些，事实上仍要到了写完全部内容才能说大功告成。

如果作品总不能最终完成，作者就只能承受挫折，不厌其烦地写稿改稿。他必须承认，作品不可能完美无瑕，所以时间可以说永远都不够。如果写出来的文章或书本从没人读，那就等于糟蹋了进入专业领域的机会。趁早别干了。

推 荐 阅 读

如果读过你草稿的人都有一致的修改意见，而且意见都与你的本意相违，那就意味着你有一个正确表达的技术性问题，如对复杂的数据分析使他们产生了误解。这也许还意味着你有不同的写作风格，但这不应是你不采纳意见的借口（见本书第26章）。

《APA出版手册》（APA, 1994）和《文体要素》（Strunk & White, 1979）都以许多事例清楚说明，怎样的写作方式是正确的，怎样是错误的。当你们在文体的问题上有抵触时，就要参阅这两本书（见本书第21章）。

其他有用的资源包括Gelfand和Walker的《掌握APA文体》（*Mastering APA Style*, 1984）、Runkel和Runkel的《社会科学作家和学生的实用指南》（*Guide to Usage for Writers and Students in the Social Sciences*, 1984），以及Becker的《社会科学家的写作》（*Writing for Social Scientists*, 1986）。Swan和Gopen（1990）讨论了科学写作中的“科学”，而Sternberg（1992）则讲述了写出好文章的一些技巧。通常，我都推荐有抱负的写作者带着

学习有效策略的目的, 去广泛阅读好文章或好书籍。

参 考 文 献

- American Psychological Association. (1994). *Publication manual of the American Psychological Association* (4th ed.). Washington, DC: Author.
- Bandura, A. (1986). *The social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.
- Beck, A. T. (1976). *Cognitive therapy and the emotional disorders*. New York: Harper & Row.
- Becker, H. S. (1986). *Writing for social scientists: How to start and finish your thesis, book, or article*. Chicago: University of Chicago Press.
- Gelfand, H., & Walker, C. J. (1994). *Mastering APA style*. Washington, DC: American Psychological Association.
- Merton, R. K. (1968). The Matthew effect in science: The reward and communication systems of science are considered. *Science*, 159, 56-63.
- Poincaré, H. (1913). *The foundations of science*. New York: Science Press.
- Runkel, P. J., & Runkel, M. (1984). *A guide to usage for writers and students in the social sciences*. Totowa, NJ: Rowan & Allanheld.
- Silver, M., & Sabini, J. (1981). Procrastinating. *Journal for the Theory of Social Behaviour*, 11, 207-221.
- Sternberg, R. J. (1992). How to win acceptances by psychology journals: 21 tips for better writing. *American Psychological Society Observer*, 5(5), 12.
- Strunk, W., & White, E. B. (1979). *The elements of style* (3rd ed.). New York: Macmillan.
- Swan, J. A., & Gopen, G. D. (1990). The science of scientific writing. *American Scientist*, 78, 550-558.



第 23 章

修改研究报告

Donna K. Nagata

Steven J. Trierweiler

多数作者在将最后的修改稿邮给杂志编辑之后都会感到无比的放松和满足。此时，你已经在文章的构思、分析和写作上花费了大量的精力（分别见目录中第2、4、5部分中这些方面的内容）。现在到了审稿阶段。美国心理学会（APA）《出版手册》（*Publication Manual*, APA, 1994）对这个过程进行了描述，并提供了一个有用的流程图来描述这个过程的每个步骤（p. 303）。在投稿几个月之后，你将收到编辑的反馈，包括是否接收你的文章，以及审稿人对你文章的修改意见（APA《出版手册》指出，尽管并不要求编辑提供审稿者意见，但他们通常会这样做）。本章将讨论在投稿之后要如何修改研究报告。首先，我们将讨论杂志审稿人的意见以及作者应如何对这些意见做出回答；其次，本章说明了在改写过程中的不同决策点；最后，我们给出一些针对二次投稿的具体建议。

尽管本章是针对那些打算在修改文章之后向同一杂志再次投稿的作者，但值得注意的是，这并不是唯一的选择。在仔细阅读评审意见及建议之后，你可能会觉得，审稿者要求的修改并不合理，并决定不再投向同一个刊物。你的文章与刊物之间的“拟合度”可能不太高，你可能希望将你的文章投往另一家杂志（但值得注意的是，即使做出这种决定，你可能仍然需要修改文章）。因此，本章也将讨论如何评估文章和杂志之间拟合度的若干问题。

审稿者是什么人

同行研究者互相评审是一个确保研究发现合法性、作者解释和结论总体水平的主要机制。若干著作讨论了这个编辑及评审过程（例如，APA, 1994; Ceci & Peters, 1994; Cofer, 1985; Daft, 1985; Roediger, 1987）。这是学术研究中的一个关键部分，因为一门科学的质量既依赖于科学研究本身的方法，又依赖于审稿人评审的严格性和准确性。

一般来说,编辑选定的审稿人是你文章所涉及领域内的专家。在撰写一篇文章的整个过程中,不要忘记评审你文章的人是在该领域内相当知名的人,而且可能是本国或世界上任何地方最优秀的专家。他们的评审工作通常是匿名的和完全自愿的。他们通常不会轻视自己的责任,尽管每次对评审工作投入的精力会有所不同。当这个评审系统正常发挥作用时,其传统性及整体性是值得敬畏的。主要的杂志在一年中会收到上百份投稿,但只有相对很少的一些被接受并真正发表。参见《心理学杂志:作者使用的资源列表》(*Journals in Psychology: A Resource Listing for Authors*, APA, 1993),可得悉一些杂志的具体编辑要求,以及这些杂志每年发表文章的数目。高投稿数与低发表率之间的对比意味着,如果你得到了一个彻底的评审,不论它是正性的还是负性的,一定有人在评审你文章的过程中做了大量的工作。

在你文章发表的过程中,要记住编辑和审稿人是站在一个整体系统的高度,而你却通常完全沉迷于自己的研究。这意味着评审稿件的人会接触大量的信息,并且会受到一些影响,如相关领域内的其他稿件。同时,他们对该领域的总体印象也是影响评审的重要因素;而你却接触不到这些信息。这个问题对于新手来说尤其重要。为了使科学达成普遍的、一致性意见,作者必须尽力弄清楚他们的工作与整体系统的拟合度。你与这个系统的关系处理得越好,你在发表文章的过程中与别人交流的机会就越大。

收到评审意见

来自杂志编辑的通知一般包括一封编辑的附信,它是编辑根据对评审者意见的总结和解释而对接受、拒收或文章修改意见提出的建议。附信的长度和详细程度变化很大,而且多数会重新阐述评审者的意见(在某些情况下,编辑也是一个评审者;见Fiske & Fogg, 1990)。附信通常还包括两个或更多阅读你文章的评审者的意见。编辑的附信很重要,它通常会让你知道从何入手重新修改稿件以及向该杂志再次投稿是否可行等。它也会强调在改写过程中需要解决的主要问题。一些杂志也提供一些信息,比如杂志重新考虑该文章之前,你需要再作多大的修改。这能帮助你判断重新投稿所需的工作量。

害怕稿件被拒收是可以理解的,研究者会关心个别审稿者的关键意见(Fiske & Fogg, 1990)。这些意见将在你决定是否再次投稿或将文章投往其他地方时起到重要作用。它们也指导作者如何修改文稿,因为在多数情况下,编辑在做出是否接受一篇文章的最终决定时,会要求作者对文章做出修改。首次阅读评审意见时,你的反应肯定带有情绪色彩:作为一个作者,从你的角度来说,你可能会觉得评审意见中的批评是误导性的、不友好的或不正确的;而实际上,评审意见也可能是友好的和公正的,促使你注意到文章中被忽略掉的一些方面。例如,一个评审者可能会提供具体的参考文献,来扩展你作特殊分析、构思或解释的思路。这里要注意两点:首先,评审者在反馈中的语气会

有所不同，一些人会直接进行批评，而其他人可能会以一种更积极的方式提出相同的问题，评审意见的长度和详细程度也有变化；第二，正如Fiske和Fogg（1990）提出的，典型的审稿者只花极少的时间提供正面反馈。所以，作者对此应该有心理准备。

初读某审稿人的批评时，你可能会感到失败和失望。这是可以理解的，因为你没有办法立刻对审稿人的评语做出反应或驳斥。这种情感反应可能并不是坏事，它可以激发你修改文章及解决所提出问题的热情。任何修改的目的应该是一种展示自己研究的强烈愿望。所以，最开始处理审稿意见的策略是先通读它们，然后在修改文章之前将它们放置一段时间，这有助于使你从一开始的情感反应中解脱出来。与开始相比，通常在重读评审意见时，你会觉得评审意见的负面性少一些，并能意识到自己忽略了评审意见中一些正面的信息。这些想法可以给你带来修改文章的信心及效率。

使用一种具体的、逐步的、针对问题的策略，可以帮助你控制由评审意见而产生的情绪，增加效能感。首要的工作应该包括以下几点：（a）评估审稿者提出的问题有多严重；（b）权衡通过修改能在多大程度上解决它们；（c）判断修改是不是一个合适的途径，如果是就这样做；（d）做一个修改计划，包括时间安排。

如果在一次改稿过程中需要解决多个批评意见，作者在一开始可以只是简单地在一张纸的一侧列出对这些批评的反应，在另一侧列出解决它们的策略。如此做时，要仔细考虑以下的问题及策略。研究中的主要观点是否受到了质疑？如果是这样，能否解决？或者，审稿者是否在某种程度上误解了一些东西？如果是这样，为什么会发生这样的事情？永远不要认为你的写作方式已经足够清晰了，有时候在写作方式上的微小变化可以让审稿者弄清楚他们在审稿时无法理解的观点。仔细考虑你的文章中被审稿者评议的每个观点。但是，审稿者没有评议某些部分，并不意味着这些部分不需要改进。想像自己正在与审稿者对话，对他/她可能说的话进行反应，并且预想一下他们在处理你的修改稿时的可能反应。解决每一个审稿意见，每次一个。不要忽视或轻视任何意见，而是客观地把它当做是从你自己的角度出发提出的观点。因为时间和交流上的限制，审稿者并不是总能彻底讲明白他们觉得什么地方不对。一些人会直接告诉你；其他人可能通过关注一个相对微小的问题来表达观点，而在语气上似乎小题大做。尽量读懂审稿者意见的根源所在。与审稿者相比，你在自己的文章上永远是专家。如果存在某种不同意见，只有直接解决它，才能成功达到修改的目的。要认识到，任务的最终目的是要让审稿者以及编辑确信他们的问题已经被充分解决。要始终从一个平衡并且清晰的角度来工作；审稿者是你面对听众的窗口，甚至当他们的反馈模糊、不确定时，对它们的关注也可以在很大程度上改善你的研究表述。

不要期望多个审稿者会在他们的审稿意见上达成一致。Fiske和Fogg（1990）指出，审稿人评审意见之间的相关一般在0.3左右。他们对APA出版的12种杂志编辑收到的对153篇投稿的400多篇评审意见进行研究，发现“一个评审者对一篇文章的评审很少与其他审

稿者一致”（p.593）。这意味着在修改文章之前你应该仔细考虑所提问题的全部范围。

虽然在某些问题上缺少一致的意见，但文章某些方面的弱点通常会在审稿者意见中提及。Fiske和Fogg（1990）检查了他们抽样的审稿者所报告的10种弱点所在：在引入自己观点之前对理论框架的描述，理论框架与自己观点之间的衔接，设计，程序，测量，统计分析，结果，解释和结论，格式和写作方式以及其他一些弱点。在这10点里面，最常提及的是“解释和结论”，其后依次是在引入自己观点之前对理论框架的描述、结果、程序和设计。也可参见Daft（1985）名为《为什么我建议拒收你的稿件以及你能为之做什么》一文，他列出了他自己审查的100多篇稿件中通常存在的问题。Fiske和Fogg（1990）给出的另外一个参考资料是Smigel和Ross（1970）从编辑角度总结的接受或拒绝稿件的原因。

修改研究报告时通常要涉及的部分

一篇文章中需要修改的具体部分当然会有差异。审稿者可能要求作者扩展文章引言，或增加解释与结论部分的覆盖面。编辑/审稿者也可能认为研究中需要做额外的或不同的统计分析。在这种情况下，修改过程中不仅需要进行更多分析，而且需要修改结果和解释/结论部分，以对这些意见作出反应。与Fiske和Fogg（1990）讨论的稿件中的10种弱点分类一致，我们将审稿者意见分为5个需要注意的部分：（a）理论；（b）方法；（c）数据分析；（d）解释；（e）与刊物的契合度。在这部分，我们将概述这些批评意见与如何受制于研究报告的结构，并讨论对文章修改的含义。如果评审者的意见不是特定性的，作者就要特别注意评审者没有提及的部分是否要加强，以免将来评审者再提出额外的意见。

修改出版扉页

一篇稿件的扉页会形成评审者对文章以及它所描述的研究的期望。这里一个很小的疏忽或不明确都有可能使评审者对文章作出完全不同的反应。同样，一个小小的变化有时能够完全消除评审者的担心。永远不要低估文章题目和摘要的迷惑性，有时它们能使评审者预期一些你没有包含在文章中的东西。你的文章题目包含一些与你报告的中心主题直接相关的词或短语吗？摘要与你在文章的主体部分中详述的内容一致吗？如果不是，那么考虑改变它们或修改稿件的一些部分，以加强扉页和文章主体间的逻辑联系。

修改引言部分

引言包括三个主要的成分：关于研究问题的陈述，文献综述，对基本理论框架的详述。Fiske和Fogg（1990）指出，这些成分中的任何一个或者全部都可能遭到评审者的质

疑。这些不同的成分之间通常都有很强的关联。与扉页一样，问题陈述可能会导致评审者对接下来出现的内容产生期望。讨论的文献应该直接聚焦于研究的问题，给读者提供关于这一问题的研究状态。作为相关领域中的专家，评审者可能会注意你对文献陈述的完全性、适当性以及准确性。有时，产生的问题可以通过增加、甚至是减少参考文献来完成。然而，评审者经常关心的是对理论框架更为详尽的论述：如果作者过分注意其他人已经提出了什么观点，而缺乏对自己研究合理性的论证，文章往往就会显得含混不清。对研究报告来说，好的文献综述应该主要集中于手边的问题，并只提供足够与当前研究直接相关的细节，指导当前研究的理论框架应该尽可能具体、明确。对它们的评价，通常是按照适合度、充分度、对随后方法和结果以及它们把文献综述内容与你研究发现联系起来的清晰度和逻辑性。一个设计良好的图形能把理论的关键结构及其相互关系以视觉的形式展现出来。鉴于它在总结复杂信息方面的巨大作用，我们推荐作者在任何可行的时候采用这种形式——定性或定量信息。（Tufte, 1990）。

修改方法部分

如果评审人发现设计或研究的实施存在缺陷，比如出现了混淆变量，重新收集数据可能会是唯一的修正办法。另外，评审者会关心方法阐述的清晰性，可能会注意诸如被试样本、研究设计、测量以及程序等问题。在一个设计很好的研究和成功的研究报告中，这些元素将会很清楚地表述，并且与整个研究的目标有明显的联系。这部分写作的目的是得到评审者在阅读时的点头赞同，很少或没有奇异难解的地方。如果你未曾报告诸如信度系数等数据，好的评审者和好的读者就会注意到，因此，对它们进行全面的、无保留的报告十分必要。甚至一些有问题的研究情境，如果作者处理适当，评审者也能有很好的感觉。在修改时，注意并修正那些可能会引起评审者怀疑的部分，让他们感到作者已完全陈述其所做的工作；对方法部分的怀疑可能很快会蔓延到稿件的其他方面，极大地阻碍稿件的接受与发表。当然，评审者也可能会发现研究中的一些设计缺陷或弱点，这些问题有时在修改时并不容易克服。如果这样，你应该对研究进行重新考查，尽量确立能够得到支持的结论。如果解释和结论的水平与研究的优势和弱点相一致，发表这些结果还是可能的。评审者也可能会给出明确说法，如果想使文章达到发表水平，还需要做哪些工作。

修改结果部分

对许多研究，尽管科学家们不愿意承认，但结果总体上来说比较难以报告。新手们经常忽视展示研究结果时的细微之处，尤其是做到既一目了然，又使结果报告集中于感兴趣的问题。巧妙之处在于找到结果部分合适的聚焦水平——不要太细致，也不要太概括——在组织行文时，要做到既把读者带你研究的最基础的水平，又把他们带回来，

准备接受你的结论。结果的呈现方式和呈现顺序都应体现研究问题、研究设计以及主要结果之间的逻辑联系。非主要却很重要的信息可以留到文章的讨论部分再展示。报告结果的全面性很重要，但评审者不会欣赏不必要的、烦人的细节。报告统计结果时，精确和全面非常重要，但是不要落到过分依赖统计复杂性来加深评审者印象的陷阱里。研究应该新颖、深刻，但评审者最终想知道的是研究的数据怎样说明了研究的问题。统计检验阐述了结果中观察到的差异在多大程度上是随机产生的；这个几率一旦说清楚，它对你的表述就没有多大用了（Cohen, 1994）。你研究的本质是在于数据的结构，比如在你设计的条件组均数之间差异的大小。在表达统计结果的同时，报告描述性和总结性的信息是没有错的——如果不需要的话，随时可以删去——但如果你不这样做，就可能会犯错误。如果结果部分需要修改，应该按数据的重要性加以梳理，包括对图表进行编辑和补充，这样最可能成功。同样，在陈述时有一个逻辑顺序，这会有助于读者快速掌握研究要检验什么以及如何检验。这里组织得越紧密，评审者就越不可能提出问题。

修改讨论部分

讨论部分也是很难写的。在这里评审者通常会发现，在第一次提交的稿件中总有些地方不恰当。一个好的讨论，其目标就是揭示研究的主要问题，得到对理论和研究的启示，并以与研究的优点和局限相一致的方式得出结论。评审者们会在讨论中寻找所有这些，但他们不期望花费太多的精力来理解它们是怎样综合起来的——因为这是作者的工作。不幸的是，在我们作为评审者的经验中，经常会发现一些原本可以写得很有力度的报告，在讨论部分却不了了之，或者对研究的理论内涵只给予微薄的探讨。这通常会使人觉得，作者没有对他们工作的重要意义予以认真的思考。或者更糟，他们看起来好像在等评审者批评之后才发表个人意见。虽然这可能是个促使文章发表的有效策略，但似乎有点不择手段，留给评审者去思考论文需要怎样解释极其不恰当的。因此，讨论部分的典型修改意见是考虑如何扩展对结果的解释，以便把它们放在一个更大的理论框架之中；或者修正其过强或过弱的结论，以适应研究设计的局限性；或者是将结果与当前其他理论上、经验上有关的领域联系起来。因为讨论是解释性的，所以对它们的写作和修改没有简单的规则。然而要记住的是，解释应该在围绕数据与更广泛的涵义之间保持一定的平衡。

一个组织良好的讨论部分会引发一些未来的研究，所以一定要仔细考虑，你想让读者在论文得出结论时得到什么。你把这些想法说得越明白、越深刻，那么评审者的反应就会越积极。此外，如果评审者发现了他不同意的地方，你在捍卫（或修改）你的观点而与评审者进行的对话中，可能会极大地改善你的论文并提高对研究的理解。记住，科学的一个重要理想就是围绕一个研究领域建立一致的意见，并且为这一致意见建立实证基础。在向这一目标努力的过程中，你会注意到，一些解释结果的具体办法是很少有争

议的。然而，问题来自于如何把这些基本方法与高级的理论解释相联系（如，Cronbach & Meehl, 1955）。这是一种艺术，学习和练习这种艺术的最佳途径可能就是根据评审者的意见而对科学报告进行修改。

评估文章与杂志的匹配度

在许多情况下，你投稿刊物就是最适合你发表该项研究报告的刊物，但要注意的是，评审者会考虑一篇手稿是否与刊物通常的质量和形式相契合。《出版手册》（APA, 1994）描述了APA旗下刊物的宗旨以及出版的文章种类。许多描述表明了接受文章发表的主要标准以及合适的研究领域。作者可以参考《心理学的杂志：作者使用的资源列表》（APA, 1993），以及《行为科学中杂志的写作指导》（*Author's Guide to Journals in the Behavioral Sciences*, Wang, 1989）。然而，APA《出版手册》建议作者多参考相关杂志最近的宗旨与规则陈述，因为任何改变都会体现在其中。

评审者将判定你的研究在多大程度会引起该杂志读者的兴趣，在多大程度上能提高该领域的知识。很明显，这是一种比较性的相对性的判断，它既依赖于所有提交给该杂志的文章的总体质量，也依赖于某篇特定稿件的独特长处。杂志在搜寻最新研究工作的程度上可能会有差别。根据你从评审者那里得到的反馈，你可能利用这次修改的机会，重新评价是否可以把你的稿件提交到最好的杂志上。什么是“最好”的杂志呢？许多人认为，发表在APA杂志上是最让人满足的。然而，一个APA杂志是否为某个具体工作的最好出版途径、是否拥有最合适的读者群，这完全是你的个人决定。因此你可能也想考察一下其他杂志，但不管怎么说，应该是一个有同行评审的杂志，而不是一个没有评审标准的杂志。

在编辑的回信中，他们对文章修改有一个暂时的判断；毕竟，他们也不知道你能修改到什么程度。然而，一般来说，建议作者“修改并再次提交”是一个积极的反馈，至少表明你的研究文章可能会被该杂志采用。你应该记住，一旦写出了一篇好的稿件，你的兴趣就应该在你的文章能在什么杂志上得到恰当的曝光率。利用修改的机会，仔细地重新考虑这一问题。

在进行修改时要考虑的其他问题

上面我们已经在很大程度上讨论了修改一篇研究稿件时的关键信息——来自杂志编辑和评审者的反馈。如果一个人想把工作发表在一个同行评议的杂志上，对这些反馈信息给予关注是非常重要的。然而，修改过程要比这广泛得多。字面上，修改就是“再看看”一个人的工作。根据评审者的反馈以及他（或她）对文章的重新阅读，作者可能会想增加额外的新内容。在这种情况下，评审者的意见可以提供给作者论文以外的观点。作者在应对评审者的批评的同时，也获得了一个考察自己工作的新视角。同样，在把文

章提交给杂志之前，先由同事读一读修改稿也是很有益处的，因为作者可能在写作过程中过于投入而无法发现文章中的缺点。

事实上，重写一篇研究稿件的挑战是，如何在引入对各种修改的同时保持一个逻辑一致的、贯通的写作风格。虽然文字处理软件中“粘贴”的功能很有用，但其可能会使稿件的整体思路发生很大的变化。在修改阶段，作者应该小心地回顾自己的论文，注意写法、语法以及标点。Fiske和Fogg（1990）指出：“评审者在文章的表述、解释与描述中看到的问题，并不少于其在实际研究中碰到的问题（p.592）。”很明显，在浏览一篇再次提交的稿件时，编辑和评审者可能会同样注意这一问题或者对它们更为关注。

在撰写手稿以及再一次修改时，一定要注意整体的统一性和写作的紧凑性。这些好的写作品质可能比较难以达到，但是随着修改，会变得越来越明显。它们给读者一种每件事情都以清晰、连贯的方式结合在一起的感觉；读者被文章的进程引领着，即使在最复杂的部分，他们依然会觉得相对轻松。这不是说你不能写复杂的问题，而是要努力思考读者阅读时的思路，想办法帮助他们更方便地阅读。核心问题是效率。一些认真挑选的过渡或小标题可以起到非凡的作用，明智的作者会特别注重稿件的这一方面，因为评审者肯定会对他们的努力做出积极的反应。

一旦你感觉修改过的稿件已经可以再次提交了，那么在邮寄给杂志编辑前你应该为重写的稿件准备一封附信。除了说明这篇文章是再次提交的，通常还要描述你针对编辑以及评审者的意见在哪些方面做了修改。如果存在你决定不改变或不作应答的问题，你应该清楚说明你这样做的理由。这会提醒编辑一个事实，即你并没有忽略评审者的建议，并且你为自己的修改提供一个合理的解释。总之，这样的一封信从总体上说明，修改过的稿件已经考虑了先前主要的修改意见（请参考本书第26章，里面有一些应对编辑和评审者的其他信息）。

最后还有两个想法。如果评审者已经为你做了工作，无论他们的工作看起来是容易的还是困难的、公正的还是误导的，你重新提交的稿件很可能会比原来的要好。评审者的努力值得在论文的致谢中提到。另外，我们把修改看作是发表过程中一个平常的、起作用的步骤。但它不是没有缺陷：例如，更加清晰的修改稿有时会引发评审者提出一些不可预料的新问题。编辑一般不会把作者放到双重的危险之中，但是他们也不会忽略在评审过程中的重大问题。对作者来说，应该认识到，这只是达成一致意见过程的一部分，不管这个过程多艰难。都是一个学习的过程，可以学到关于你的研究应该如何被概念化、如何在评审过程中与同行交流、关于你自己和工作的一些新知识。这样的态度有助于你接受发表过程中的必要规则，并可使修改手稿的过程变得更为有趣。在任何情况下，把你的结论呈现给同行是科学实践的重要过程，尽管这一过程不会让你喜欢上修改自己的作品。与最初提交稿件一样，当修改完成并感到稿件已被比较恰当地改进之后，作者会产生一种解脱的感觉。然而，应该记得，虽然通过修改可以提高稿件被接受的可

能性,但修改并不能保证稿件最终一定被接受。Fiske 和Fogg (1990)指出,只有2%的最初投稿会被接受(Eichorn & Vandenbos, 1985),修改稿件中也只有20%~40%会被接受。评审和发表的过程是一个不完美的艺术形式(Kupfersmid, 1988; Sussman, 1993),评审者的可信度也受质疑(Peters & Ceci, 1982; Spencer, Hartnett, & Mahoney, 1986)。但是,同行评审是科学的标准方法,它不会很快发生,也不会很快改变。依靠幸运和勤奋,你的努力会得到回报,你的稿件会被接受,你可以骄傲地看到你的工作结果正在被印刷出版。

推 荐 阅 读

APA《出版手册》(APA, 1994)对本章中提到的很多问题有一个很好的概述,包括稿件准备的细节、编辑和评审过程、一般的写作指导。其他涉及编辑和评审过程的书籍包括Cumming和Frost的《组织科学中的出版》(*Publishing in the Organizational Sciences*, 1985)、Jackson和Rushton的《科学的卓越:起源和评价》(*Scientific Excellence: Origins and Assessment*, 1987)、DeBakey的《科学日记:编辑规则和实践》(*The Scientific Journal: Editorial Policies and Practices*, 1976)。对这些问题的新看法,可参考Sussman (1993)的文章《学术出版中的卡律布迪斯》。还有几本关于图、表以及信息视觉呈现的优秀书籍,包括Tufte的《预想信息》(*Envisioning Information*, 1990)和《定量信息的视觉呈现》(*The Visual Display of Quantitative Information*, 1983)。其他的阅读材料有Kosslyn的《图形设计元素》(*Elements of graph Design*, 1994)、Wilkinson、Hill和Vang的《制图法》(*Graphics*, 1992)。

参 考 文 献

- American Psychological Association. (1993). *Journals in psychology: A resource listing for authors*. Washington, DC: Author.
- American Psychological Association. (1994). *Publication manual of the American Psychological Association* (4th ed.). Washington, DC: Author.
- Ceci, S. J., & Peters, D. (1984). How blind is blind review? *American Psychologist*, 39, 1491-1494.
- Cofer, C. N. (1985). Some reactions to manuscript review from a questionnaire study. *Behavioral and Brain Sciences*, 8, 745-746.
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 49, 997-1003.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Cummings, L. L., & Frost, P. J. (Eds.). (1985). *Publishing in the organizational sciences*. Homewood, IL: Richard D. Irwin.
- Daft, R. L. (1985). Why I recommended that your manuscript be rejected and what you can do about it. In L. L. Cummings & P. J. Frost (Eds.), *Publishing in the organizational sciences* (pp. 193-204). Homewood, IL: Richard D. Irwin.
- DeBakey, L. (1976). *The scientific journal: Editorial policies and practices*. St. Louis, MO: C. V. Mosby.
- Eichorn, D. H., & VandenBos, G. R. (1985). Dissemination of scientific and professional knowledge: Journal publication within the APA. *American Psychologist*, 40, 1309-1316.

- Fiske, D. W., & Fogg, L. (1990). But the reviewers are making different criticisms of my paper! Diversity and uniqueness in reviewer comments. *American Psychologist*, 45, 591-598.
- Jackson, D. N., & Rushton, J. P. (Eds.). (1987). *Scientific excellence: Origins and assessment*. Newbury Park, CA: Sage.
- Kosslyn, S. M. (1994). *Elements of graph design*. New York: W. H. Freeman.
- Kupfersmid, J. (1988). Improving what is published: A model in search of an editor. *American Psychologist*, 43, 635-642.
- Peters, D., & Ceci, S. J. (1982). Peer-review practices of psychological journals: The fate of published articles submitted again. *Behavioral and Brain Sciences*, 5, 187-255.
- Peters, D., & Ceci, S. J. (1985). Peer review: Beauty is in the eye of the beholder. *Behavioral and Brain Sciences*, 8, 747-750.
- Roediger, H. L., III. (1987). The role of journal editors in the scientific process. In D. N. Jackson & J. P. Rushton (Eds.), *Scientific excellence: Origins and assessment* (pp. 222-252). Newbury Park, CA: Sage.
- Smigel, E. D., & Ross, H. L. (1970). Factors in the editorial decision. *American Sociologist*, 5, 19-21.
- Spencer, N. J., Hartnett, J., & Mahoney, J. (1986). Problems with review in the standard editorial practice. *Journal of Social Behavior and Personality*, 1, 21-36.
- Sussman, M. B. (1993). The Charybdis of publishing in academia. *Marriage and Family Review*, 18, 161-169.
- Tufte, E. R. (1983). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.
- Tufte, E. R. (1990). *Envisioning information*. Cheshire, CT: Graphics Press.
- Wang, A. Y. (1989). *Author's guide to journals in the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.
- Wilkinson, L., Hill, M., & Vang, E. (1992). *Graphics*. Evanston, IL: Systat Inc.



第 24 章

应对杂志编辑和评审人

Samuel H. Osipow

科学家的一部分责任是与大家共享他们的成果。因此，当一个研究项目涉及的所有工作完成之后，应该将它们发表出来。但在达成以上目的的过程中，有许多潜在的障碍，很多障碍来自期刊编辑和审稿者。下面给出的一些基本程序与方法，可以减少发表文章过程中的挫败感，这些程序甚至有提高文章发表的可能性。

准 备 文 稿

第一步在研究过程中就已经开始了。一个好的（即重要并有趣的）想法和一个可靠的实验设计是研究开始阶段的必要前提（见Wampold，本书第5章）。然后，数据一定要基于一个合适的、足够大的样本，且适合检验你的假说。这些假说以及它们的来源——理论或他人的经验发现——应该很明确地在文中进行阐述。文稿应该有一个全面的文献综述，阐明该研究的独特之处，证实该研究不是一个严格的复制（见Stockdale 和 Kenny，本书第3章）。最后是使用合适的统计方法，对数据进行恰当的分析，产生一定的解释。换句话说，没有什么能比好的研究结果更能确保文章得以发表了（见Dollinger & Dilalla，本书第13章；Levy & Steelman，本书第17章）。

选择一个发表途径

研究是否出色固然重要，但对发表来说还不足够。当你的研究得出结果并且你的研究报告已经写好之后，下一个任务就是为你的工作确认合适的发表途径。最容易的方法就是阅读或者浏览几期你认为可能合适的杂志。许多杂志的目标和限定陈述在杂志刊头。仔细阅读这些陈述，至少在一个初级水平上让自己相信，这个杂志可以作为你发表自己研究的一个途径。你也要看一看最近几期的文章目录，以了解最近什么样的文章才

被认为是合适的。如果你仍然不确定，那么起草一封问询信，大致总结了你的研究（在一页纸内），咨询一下编辑。大多数编辑会对这样的问询做出反应，因为它可以节省编辑很多评审时间。收到一个很显然从来没有读过该杂志的人的稿件，是编辑最烦恼的事情之一。如果你以前看过该刊物，那就会了解一些情况，不会贸然提交稿件。

在考虑杂志的合适性时，也要保证你看了最近出版的该杂志。即使在刊头部分的目标陈述在几年之内都是不变的，但杂志所偏重的主题会不断变化。对杂志来说，不同主题的重要性可能会增加或下降，不停地“漂移”。把稿件邮给当前的编辑也很重要，所以要仔细检查一下稿件邮寄的名字和地址。

在阅读刊头的陈述时，需要特别关注几项重要的技术信息，如不遵守就会浪费你的时间和编辑的好感。要搞清楚一些简单的事情，比如要提交多少份、稿件是否要准备好“盲”审以及稿件的风格和格式要求。“盲”审允许作者对评审者匿名，这对一些年轻的研究者有好处，可以让你的稿件根据本身的优点而不是根据作者的名声被评审。

在这一阶段，要用合适的格式来准备你的稿件。这里重要的一点是避免对自己的用词过于执着。通常提交出版的稿件是基于学位论文的研究。如果是这种情况，编辑经常提出的意见可能是稿件太长、提供的文献综述过于广泛、对结果和讨论过于仔细等。你一定要学会很好地使用“蓝笔”，把文章删减到杂志要求的长度，同时还要保留重要的方法、细节和结果，以便那些对你的领域不太熟悉的读者还可以理解你的文章。你可能对自己的领域过于熟悉了，以至你并不总能看到表达中存在的跳跃之处。

表和图的呈现要按照顺序。美国心理学会《出版手册》（1994）中详细描述了表和图的格式，一定要认真遵循。然而，需要强调的是，在决定什么内容应该以图表方式呈现时，作者应该做出良好的判断。如果信息用文字的形式很容易说明，那么表和图可能就不必要了。

在你写完一篇你认为不错的草稿后，把它给同事或教授看看，征求批评意见。受到批评尽管可能带来一定的痛苦，但这些反馈所提供的价值远远补偿了你的痛苦。一个“无知的”（对你的文章）读者能够察觉出你作为作者会因对这篇文章和这一领域的熟悉而忽略掉的漏洞。因此，外行的读者可能会找到表达不清的句子，如模糊的指代、混乱的组织、信息连贯上的缺陷、令人迷惑的图表以及其他。记住，从一个有同情心的读者那里得到这些批评要比从编辑或评审者那里得来好得多，因为后者的话语可能会更为苛刻。

编辑的一些典型反应

在你确信稿件已被准确打印，表格和图已被很好呈现，你就可以把你的稿件邮寄给编辑进行评审，然后就开始了漫长的等待。大多数心理学杂志在两个月内给作者返回评

审意见，但是有时因为评审者的耽搁或编辑工作的众多，这个过程可能需要花费更多的时间。如果你在三个月后还没有收到回音，询问一下稿件的状态是很合理的。

你最后得到了编辑的回音，结果可能会有多种，每种结果都会引发作者独特的反应。最好的结果当然是编辑说这篇文章很好，可以接受。为了告诉读者这一结果是多么的难得，请看以下例子。根据我作为杂志编辑的经验，我大概为接近3000篇稿件做了编辑决定，只有三次我接受了作者第一次提交的原稿，其中两次都来自于同一作者。

当然，最坏的结果就是彻底拒绝。作为作者，你可能会觉得感情受到伤害。这种情感反应是可以想见的。你的第一反应是震惊：我做了这么多工作，这是为什么？然后是愤怒：这些评审和编辑都是傻瓜，或者他们对我或我研究的问题怀有偏见。这些阶段在不同人的身上会持续不同的时间。一个经验丰富的作者会经历时间较短的、不太极端的反应，因为这样的作者对自己发表专业文章的能力有充分的自信，可以理性地对编辑的决定作出评估。在愤怒阶段，不要修改稿件或给编辑写信和打电话，草率的决定可能会使事情变得更糟。

在头两个阶段过去后，一个更为理性的反应开始出现。作者能够阅读评论，并更为客观地评估它们。有时你会逐渐明了，论文投到另一家杂志会更好。但在很多时候，评论可能会很好地指导作者发现适于发表的其他表达方式，或者发现可用其他研究修正缺点。有时，一个作者可能逐渐达到另一个阶段，此时他或她想感谢编辑和评审者拒绝了自己的论文，因为这种拒绝避免了作者因发表带有重大缺陷的文章而在全体同行面前产生的尴尬。

无论结果如何，要记住，许多心理学杂志拒绝85%~90%的稿件，所以不要对一个拒绝做出过分的反应。如果你相信你的文章尽管被拒绝但还有根本的价值，想一些办法来修改它并换投一家新的杂志，这使一个持之以恒的作者通常能够成功地发表文章，虽然可能不是在第一选择的杂志上。

还有两种编辑决定也经常出现。编辑可能写到“这篇稿件不能按照现在提交的形式被发表，但如果你愿意按评审者的建议作一些修正（见信里附页），我们会考虑修改后的稿件。”这种情况下的中心词是“考虑”。这里，你一定要意识到，编辑并没有做出承诺，如果你修改论文就一定会发表，他只是做了重新评审的许诺，这是编辑们经常做的事情。这种决定可能会令人沮丧，包含某种陷阱。作者可能会花费大量的时间来修改稿件，但在文章重新提交后，编辑只是把稿件送到一个或更多新的评审者那里。然后他们可能发现一些与前次不同的“缺点”。在这种情况下，修改可能会不断地进行下去。如果这种情况发生在你的身上，那么给你的论文设定修改次数可能是比较明智的。

一个较好的结果是：编辑告诉你，如果你愿意做某些评审意见建议的变化，那么文章就可以发表，然后告诉你什么必须要改变、文章的缺点在哪里。这里重要的是，在进行修改前一定要清楚地知道需要改变的地方。如果你不是很清楚文中需要作的改变，尽

管向编辑询问。

修 改 稿 件

一般情况下，当给出“修改后可以接受”的决定后，作者只有有限的时间进行修改（见Nagata & Trierweiler，本书第23章）。如果你认为编辑没有给你足够的时间，你需要花费比编辑估计的更多的时间和努力，或者你很忙而可能耽搁稿件，那么给编辑写信或打电话请求更多的时间，并说明你还需要多少时间。如果及时提出要求，而要求看起来是合理的，那么大部分的编辑都会同意延迟。

要记住的另一点是，编辑的反馈应该是具有教育意义的。如果可能的话，应该假定编辑和评审者都是在试图帮助你，站在你的一边。许多编辑（如果不是全部的话）还是把他们的工作看作是教会新手如何发表文章的过程。

不管上面描述的三种结果中的哪一种出现，编辑一般会有详细的反馈，作者应从中学会很多关于做研究或描述结果的知识。当然，教育这一阶段只有在作者的震惊和愤怒阶段过去之后才会发生。

应对苛刻的反馈

阅读稿件反馈中最难的是判断评审者评论的语气。编辑经常会把评审者的未雕琢的批评副本封在信里。作为作者，有时需要明白评审者和编辑在感情方面具有本质上的区别。不幸的是，评论通常是极端具有敌意和贬低色彩的。这些其实都很常见，他们的批评超过了真正的建设性，并且进入了人身攻击的范围。通常编辑在给出反馈时会比较圆滑，但是有时编辑也会采用那种愤怒的且贬低身份的方式。在我曾经编辑过的一本杂志（《应用和预防心理学》）中，通常要求评审者避免采用敌对的方式，且让他们明白其作用是使稿件尽可能的好。经验已经表明，不仅作者会欣赏这种方式，评审者也是如此。不幸的是，不同的杂志可能并不都使用这一方法。

一个人可能会认为审稿人一定专门针对那些被拒绝的稿件而采取这种愤怒、敌意的批评。然而，他针对的也可能是那些修改后会接受或者邀请再次提交的稿件。这一因素使得所有作者更难通过愤怒反应进入下一个建设性修改阶段。作者所能做的最好的事情就是，首先把评论先放在一边大约1~2天。评论可能在一些天里一直看起来都很差劲，但是当作者变得越来越平静时，就可以提炼出需要重点修改的地方，这时就没有必要再去重读那些只会再次激起愤怒情感的非建设性的评论了。

与编辑建议的不同意见

这时，作者能够开始根据评论中的有益之处修改稿件。正如前面提到的，当经过一段不长的时间后，评论可能看起来威胁少一点，并且作者对它们的情感反应可能逐渐减少到一个可以控制的水平。作者很可能会使用好的理由来辩驳具有一个或多个建议的修改意见。可以在再次提交的封面信件中提及这些辩论。

事实上，如何解决编辑关心的每个问题是非常有用的。这样才能使这些解释自然地排列，方便在合适的地方做出相应的反应。一些编辑实际上要求作者写这样一封信，来详述他们对于每一个需要改变的地方所进行的处理。这样一封信可以起到服务编辑和作者两方面的目的。对于编辑，这封信可以使他快速总结出需要的和已经做出的变化，使编辑能够对这一版本满意与否做出快速的决定。对于作者，这封信可以有相似的目的。它可以作为一个清单，作者可以使用它来检查是否已经完成了所有的修改。

当编辑的建议看起来显然是不正确之时，你可能会和编辑争论。这种行为需要小心，因为它会使你与编辑敌对，而且也浪费你自己的时间。然而，如果你不能接受一个主要的修改建议，并且你的观点不能容易地合并到前面提到的封面信件中去时，那么这就需要提出一个有效辩驳，最有效的办法是使用文献、已经存在的数据或者你自己的数据以及最重要的逻辑，以此尝试说服编辑接受你的观点。

如果曾遭拒绝，那么相应的问题就是什么时候挑战这个拒绝。我个人的观点是永远不要，但是有许多作者在这件事上取得过成功。想挑战的话，使用上面提到的同样的要点（文献、数据以及逻辑）来支持你的观点。你获得的最好结果很可能是由两个不同的编辑顾问来重新评审。

表24.1总结了本章中的一些关于如何来应对编辑和评审者的要点。

我无需强调有效的写作技能的重要性。这些技能甚至在开始出版你的研究成果的第一步前就是很必要的。

表24.1 应对编辑和评审者的一些要点

-
1. 做一个全面的文献综述。
 2. 确定你的研究的重要性。
 3. 准备一个良好的实验设计和合适的统计。
 4. 确定合适的出版途径。
 5. 使用正确的格式来准备稿件。
 6. 不要对编辑的反馈或决定做情绪上的反应。
 7. 确信你明确地理解了编辑所要求的修改。
 8. 写一个关于你所做的改变的解释性信件。
-

推 荐 阅 读

我们推荐读者参考下面的一些资料,以便更深入地理解本章中讨论的主题:Daft (1985)、Frost (1985)、Weick (1985)和Dorn的《专业出版的发展》(*Publishing for Professional Development*, 1985),尤其是Gelso和Osipow (1985)所著的一章。

参 考 文 献

- American Psychological Association (1994). *Publication manual of the American Psychological Association* (4th ed.). Washington, DC: Author.
- Daft, R. L. (1985). Why I recommended that your manuscript be rejected and what you can do about it. In L. L. Cummings & P. J. Frost (Eds.), *Publishing in the organizational sciences* (pp. 193-204). Homewood, IL: Richard D. Irwin.
- Dorn, F. J. (Ed.). (1985). *Publishing for professional development*. Muncie, IN: Accelerated Development.
- Frost, P. J. (1985). Responding to rejection: An author's view of the rejection of an academic manuscript. In L. L. Cummings & P. J. Frost (Eds.), *Publishing in the organizational sciences* (pp. 766-773). Homewood, IL: Richard D. Irwin.
- Gelso, C. J., & Osipow, S. H. (1985). Guidelines for effective manuscript evaluation. In F. J. Dorn (Ed.), *Publishing for professional development* (pp. 68-69). Muncie, IN: Accelerated Development.
- Weick, K. E. (1985). Editing a rejection: A case study. In L. L. Cummings & P. J. Frost (Eds.), *Publishing in the organizational sciences* (pp. 774-780). Homewood, IL: Richard D. Irwin.



第六部分

其他主题



第 25 章

协调研究团队 ——良好实验室的保持和发展

Dennis L. Molfese, Kristen L. Murray

Tina B. Martin, Cindy J. Peters

Arlene A. Tan, Leslie A. Gill

Panagiotis A. Simos

在当今复杂的科学世界中，研究者越来越多地在研究团队中工作，这样可共享资源和才智。科学领域的复杂性已发展到一个高度，个人难以具有进行广泛实验研究所需的智力和技能（更别提时间）。考虑到这些局限，组成研究团队是在一个快速发展的领域占有一席之地最为有效的方法，它提供了应对发展所需的广泛而复杂的技能。从某种意义上说，研究生教育其实是学生未来在科学界工作的缩影，即成为科学研究团队的成员。经过研究生培训，学生在该领域中掌握大量技能，这对他们将来的职业生涯很有帮助。同时，一个研究团队是否成功，在很大程度上取决于其成员的主动性和专业技能。这意味着，即使是研究团队的最新成员，也是团队成功的重要部分。本章就壮大研究团队的多种因素发表评议（也请参阅本章末尾列出的推荐读物）。

主 动 性

学生的主动性是影响研究环境的、最为有效的途径。学生所受的教育和通过研究训练所学到的东西，在很大程度上取决于他们投入的时间和精力。这个过程并非是学生单方面汲取信息，而是需要投入大量的时间和资源。学生必须主动参与并积极寻找训练的机会。他们投入的时间越多，则获益越多；反之，投入的时间越少，则收获越少。学生如能成为团队中的一员，很自然就有获取知识和经验的机会。然而，如何利用这些机会以及能从中获得什么，则取决于学生在这种环境中做了什么。

主动性是训练和团队成功的重要因素。因此，有关主动性的一般规则在此也适用。如要做一些事情，不要总等其他人去完成。如果你知道怎样做，则应立即动手，不要拖

延时间，直到完成任务。但是研究的主动性远不止于简单地执行任务，参加研究活动对你有很多益处。你学到的技能可能伴你终生，有助于你将来的职业发展。这些技能也影响到科学界对你以及你的潜力的认同。所以，你应该全身心地投入到实验室的活动中去，不能等其他人为你设定每个训练目标；相反，你必须主动去设定。如果导师未能与你碰面，讨论训练或考核你的工作，你应该主动与导师联系，获得新的会面机会。

广泛阅读。除指定的读物之外，请导师指点一些读物。养成每周浏览相关期刊的习惯，不能把学习局限在狭小的范围内，因为没有人能预知完成一个目标需要多少信息。你可以与实验室、研究所、校外专家保持联系，以得到相关的文献或就自己研究中的问题咨询他们。除了学习感兴趣的知识，以及在实验室进行活动之外，你还应利用讲座和大会，更多、更广泛地了解其他人的研究工作。当你获得新信息时，要与实验室主任以及其他同学共享，他们会赞赏你这种主动分享信息的行为。如果你在阅读和讨论中，发现了研究线索，那你要系统地组织研究设计，向实验室主任提出建议，并付诸实施。快动手实践吧！毕竟这就是科学教育的全部——获得知识和技能使你成为独立的研究者。你不可能再获得像研究生训练这样的机会了。所有的教员都是为训练你而设，你可以接近他们，利用这些专家的知识财富。科学同仁常常需要等到某一次会议才能见到，并且参加研讨会花销很大。现在，你不需要这样，只要走进系里的大厅，稍做准备，就可占用教授的时间。请利用好这些优势。很明显，对你来讲，学习的主动精神能使你获得最大的财富。这不仅是训练一般人才，而且是培训优秀人才的重要因素。当然，成功还需要其他因素，如团队领导或实验室主任的专业能力和协作能力。

团 队 领 导

团队领导可能是你的实验室主任，他已经获得了博士或硕士学位（不管是教员还是博士后），有时还是高年级研究生。有以下4个因素，决定着团队领导的工作效率以及团队的成功：（a）一系列清楚的、特定的目标；（b）清楚的预期；（c）团队和领导之间有效而持续的双向沟通以及团队内部良好的沟通；（d）共同的价值观和关注共同的问题。

首先，团队领导应对项目提出一系列明确的目标，清楚表达对团队成员的预期。如果项目资金来自拨款基金，那么基金中列出的目标就是团队的目标。但是这些目标必须非常具体，并以具体的形式告诉团队。比如，某一项目的目标是在6个月内测试40名被试，那么事先就应该做好测试安排，设定每周测试计划，即每周测试多少个被试。其次，领导必须对团队成员提出具体要求。这些要求包括怎样完成任务以及测试中的专业水准等。第三，沟通是项目成功的关键，它涉及团队领导和团队成员。团队领导必须主动、有效地与团队成员沟通。同样，团队成员间必须注意彼此沟通，并与团队领导沟

通。如果因为某些原因，沟通阻断或者变成单向的，那么就会对项目的成功构成明显的阻碍甚至威胁；如果是长期项目，那么盲目和无计划的实验将会耗费掉大量的资源。这时沟通就变得尤为重要。实验室主任和“老”学生间的灵活性也非常重要。如果能对建议和彼此关心的问题敞开心扉，那就能为更好地沟通和实验创造条件。这一点与第四个因素——共同的价值观紧密相联。如果团队成员和领导在程序和目标上有分歧，那么数据收集就会有问题。例如，某一团队成员认为某一方法并非是收集数据的最好方法，那么他就不会像其他人那样热情地坚持这方法，甚至在其工作中作别人不愿接受的变动，这就导致沟通破裂。显然，一旦发生这样的情形，其结果对项目而言是毁灭性的。研究项目的有关方面必须经过所有团队成员（领导和学生）讨论并取得一致意见，这是至关重要的。

沟 通

为能确立沟通“规则”，首先，领导必须安排每周或双周与实验室的每个新成员会面，并给予指导。这样做很重要，可以促使每一个人认识到其在实验室中的角色，并迅速发展。当然，会面最好单独进行，这让每一个成员都能与指导老师直接相处。这种毫无威胁的指导方式将有利于问题的解决。

如果团队领导的时间表排得很满，不能腾出时间来参加指导会议，可以由其他高级实验领导（如博士后）来完成。在这种情况下，两位指导者应花时间讨论学生的发展，协调、监督和指导整个试验室的成员。单独会面的重要性之一是保证了实验室工作的持续性，有助新成员形成有关实验室以及自身发展的“宏伟蓝图”。同时，这也便于成员理解自己的个人目标怎样融入到团队领导要求之中。实际上，这种指导能使学生和实验室的发展相得益彰。人们在实现与实验室相关的个人目标时，最有可能有效地实现实验室的目标。

其次，打造健康的沟通是组织实验室所有成员参加的周例会。这些会议旨在维系长期的目标、系统讨论和解决实验室的问题。这些问题可能很具体，可能是设备故障；也可能是人际上的，比如工作的分派。所有这些都需要一个公开沟通的平台。真诚而有效的问题讨论是必不可少的，这样可以防止问题变大，危害到整个研究工作。

最后，为教育实验室成员而采用羞辱的办法，不管是采用“同伴压力”还是个人私下的方式进行，即不能获得有效的沟通，也不能直接、果断地解决问题。通过挫伤别人自尊心的方式永远达不到人际关系的真正改善。

团队的研究规范

本书是为研究助手而写的，让他们通过训练而变为更称职的科学家和学者。训练过程中一个重要的部分是让所有有关的人认识到，被训练的人在此过程中可能发生错误。这里重要的一点是，学生必须认识到这些错误，并主动与项目负责人和研究团队沟通，改进程序，从而避免或减轻类似错误发生的可能性。要实现这个目标，学生和项目负责人必须认识到错误肯定会发生。同时，双方必须尽力地监督和改善其工作质量。负责人应清楚地向成员表明虽然错误是我们不愿意犯的，但又是训练过程中不可避免的。如果研究负责人对犯错误的学生进行严厉惩罚或公开羞辱，这就大错特错了。相反，负责人应该教导学生，讲解不同情况下可能发生的错误，以及怎样去避免、发现和更正这些错误。这应该以温和的方式进行，让学生校正实验程序或者检查数据资料，并鼓励学生公开讨论自身错误，特别是自己发现的那些。让学生私下向领导汇报错误，可以避免在同伴面前汇报时不必要的尴尬，这样学生和负责人能更自由地讨论问题，更可能找到纠正和避免错误的方法。会谈的结果要在实验室会上进行讨论，以此创造一个开放的学术气氛，从而使学生能公开地、自由地讨论研究过程中的积极和消极部分。学生也应认识到，虽然一个错误避免了，但另一个错误仍会出现。一旦发生错误，学生必须知道，作为研究者，他有责任去辨明、承认和更正。同时，领导必须认识到，在团队成员面前承认错误，对于团队新成员来说，都会很不舒服。研究团队高级成员甚至团队领导公开承认错误，有助于新成员克服尴尬和恐惧的心理。

研究助手的招募

负责人面临着与雇主同样的问题——怎样去评定招募的员工、怎样预测其在研究工作中的成就。不论招募研究生还是本科生，有5个因素很重要：（a）个人面试；（b）清晰的评价标准；（c）对所需完成任务的描述；（d）任务完成的时间长短；（e）多样性。

面试能让团队负责人和将来的学生面对面地讨论任务，交流彼此对任务的预期，并有机会获悉学生的背景、经历、现有技能水平和研究兴趣。负责人必须讲清楚研究项目、预期目标、对学生在项目过程中角色作用的期待。随着讨论的深入，负责人能确定学生的资历是否达到参与研究工作的基本标准。面试中，负责人要列出学生实验室每周工作时间的要求、提出工作质量的水平要求、列出可帮助学生更多了解项目的补充读物。最后，通过招募具有不同学术背景和经历的学生，负责人能确保实验室中的个体有不同的发展方向、不同的专业技能。多样化能提升讨论的水平，甚至能改善研究项目的

设计。

本科生助手

作为录取标准的一部分，研究生已达到某些要求，或符合多数实验室对研究助手的标准。但确定本科生能否加入研究计划的标准却有很大差异。一般来说，要求本科生助手的平均（GPA）成绩在4分制下至少是3.5分以上，在连续两个学期内每周至少为实验室工作10个小时。平均成绩的要求规定了对学生技能的基本要求，同时保证他们参与研究不会危害到其他课程和学业成绩。另外，对参与时间作出要求，是为了让学生在培训期结束后能参与实验室工作，保持实验室工作的连续性。这样，学生在培训期后也能对科研项目做出贡献。

研究生助手

如上所述，研究生助手经历过筛选测试，他们可能在读本科时就参与了实验室的工作，在竞争的环境下表现出色。实验室领导能读到认识该生的教授的推荐信，分享他们对学生目前能力和潜力的评价。同时，本科生进入研究生阶段，并不意味着其具有在所有研究领域都能表现出色的技能、性情和兴趣。面试为领导和学生提供一个面对面对讨论问题、增加了解的良机。

训练和可靠性检查

保持实验室测试规则、数据输入、评分、分析等的一致性，对所有实验室来说都很重要。领导和研究助手优先考虑的是训练的一致性。用书面方式列出培训目标和内容，对确定目标技能和保证培训时间都很有用。

如果训练不一致或没有定期重复，即使是已经得到培训的成员也可能产生不确定的实验结果；如果实验室的工作程序随意变化，那就会大大降低研究发现的可能性。因为存在不确定性，所以其他混淆因素也可趁机进入研究过程中，进一步降低研究效率。实验程序上的不确定性会随时间增大。结果是，人们开始无休止地修正程序，区别对待研究早期和晚期的被试，对被试的评价和判断也会产生变化。这些很显然是严重的问题，极可能损害研究工作。如果能使用实验室手册，定期地、规则地、独立地评定测试草案，并且认真地执行，就可避免这样的危险。

实验室手册

实验室手册（见本章结束处附录25.1的样本表格）是沟通目标和预期的简单方法。手

手册是每个人都能得到的信息来源，它列出了已被接受的规则和表格，从而使训练更便利。通过研读和参考手册，学生能辨明自己理解的方面或不清楚的方面。手册还能帮助他们更好地关注自己的问题。在实验室主任或者高级实验成员不在身边时，学生如存在疑问，手册还能给出一个具有权威的指导。如果当前没有制定实验室手册，学生或者实验室主任（或共同）要制定好手册，以帮助学生更好地组织任务、坚持责任和实验程序。

实验室手册应包括实验室哲学、测试目的、测试要点、测试和分析草案的详细信息，还应包括研究的道德规范、可接受的行为方式、言谈举止和服装穿着的规定。以手册的方式给出这些信息，学生就能了解实验室的行为方式以及在各种实验室活动中该做些什么。此外，实验室手册还应包括：备份所有测试工具、考虑所有测试情境和程序，并附带上一一定的解释。写手册的目的就是提供容易参考的书面记录。采用手册，能极大地为培训工作提供便利，也能把标准化的测试程序记录在案。学生要在培训之前或期间研读实验室手册。实验室手册还能让有经验的学生获益，如防止实验的不确定性、避免工作错误等。如果实验室手册中的条文适用所有成员，那么实验室成员在不确定的情况下，或在领导或高级实验成员不确定时，查阅手册可以获得帮助。实验室手册的缺点需要不断改进和更新。然而，如果指定不同的学生去更新或修正不同的部分，并且在把内容加入到手册之前，由全体实验室成员审阅通过，那就会减少此项任务所花费的时间。手册的更新给所有参与人员一种拥有感，也会获得进一步的鼓励。

实验室会议

实验室会议是加强实验室内部沟通的有效方法。这些会议提供了一个论坛，以协调研究团队、共享有关项目的信息、解决研究过程中的困难。但是，实验室会议在效率和组织方式上可以有很大差异。多长时间碰面一次？应该涉及哪些活动？会议中的讨论怎样保证围绕主题？这些只是会议所涉及的一些问题。大体上，周例会主要讨论上一周的工作，包括设备问题、培训问题以及本周的工作安排。书面的议程能把大家的注意力集中在这些问题上，并且成为会议的焦点，如果讨论偏离了议程，还可及时收回来。议程要在主任或实验室指导老师的授意下，由最有经验的学生来制定。议题或讨论的题目通常每周都一样。

一旦开展新项目，实验室增加会议时间将变得与为学生提供相关读物或书目一样重要。增加的时间主要用于讨论先前有关的工作和新项目的理论问题。因为在实验室例会（通常都有固定的时间限制）上，没有时间讨论附加的内容。所以，会议必须留出时间，系统讨论学生研究项目的结果。对学生来说，这可以让别人看到他们工作的总结，而不是一个无休止“进行中”的研究。对于刚刚接受实验室任务的新同学来说，向领导

和同学回顾并展现实验室过去的研究工作非常重要。这种回顾帮助新生（也能提醒有经验的老生）理解目前研究计划和将参加的研究的大体情况。

当然，会议需要腾出特定的时间，安排申请硕士或博士学位的学生正式答辩前在实验室演讲。这些演讲练习给予学生机会，让他们在较温和的氛围中尝试解释这些材料（即展现提纲、表格、图表等），演练表达的策略和技巧。在演练过程中提出的一些问题，可使学生在答辩前做好解决准备。当要为会议或者专业讨论会准备海报或上台陈述时，实验室可在会议外的时间组织讨论，使演讲者更好地准备陈述材料。

数 据 检 查

实验室会积累各种实验数据。累积数据时，研究人员面临许多可能危害数据完整性的问题。显然，数据分析是研究过程的一个高峰。数据保护和处理的好坏能决定一个研究项目是否成功。数据必须尽可能精确，否则在项目中所花费的时间和精力将付之东流。研究者应抽出额外时间来确保数据的质量，以保存先前工作的结果。与产出高质量的实验结果相比，检查数据所花的时间和精力只是很小的代价，特别是前期联系被试和安排测试、准备仪器等投入的时间和努力与后来的数据分析、草稿写作、研究结果的最终出版相比。

数据的形式多种多样，可以是书面的问卷，也可以是主试对被试作业的评定、有被试反应的计算机文件、被试同意参加实验的书面签字表格等。研究人员要保存每个被试的系列记录，所以他们面临重要但却非标准化的任务：汇总、保存并跟踪每个被试的材料，检查实验人员的记录材料（如测试日期、时间、条件的安排等），把数据输入电脑以进行数据分析，核查数据条目以确定输入是否准确。所有这些步骤都需要花费很多时间，需要分外小心，精确无误。研究越复杂，研究参与的人员越多，在这一阶段花费的时间就越多。实际上，在进行纵向的数据收集时，除在测试上花费的时间外，研究者70%~80%的时间都花在整理和核查数据是否完整上。

数据核查的程序在实验室手册中应有清楚描述，或在实验室全体会议上一致通过。如果没有实验室手册，这些讨论的书面记录必须保留，以备将来参考。这个过程中重要的一点就是保持警惕。一个特别有效的方法是：让一个人收集数据，而第二个人核查数据。利用计算机程序来校验数据录入非常有用。这些程序要求实验者在把数据存储到主数据库前二次录入数据，程序搜索两次录入之间的不同之处，从而让录入者解决冲突或矛盾的数据。如果可能，由第三方随机核查也很有用。永远要有数据备份，并随时作相应的更新。

由谁来收集、输入和操作数据？一般来说，最好让高年级的学生负责数据的输入和操作，因为他们在这些相关的程序上更有经验。然而，利用已输入的数据培训新生，把

现在的结果和以前已核查的结果作比较,是培训新生对程序掌握的简单有效的方法。此外,新生可以观察有经验的学生的操作,以了解这些活动具体是如何进行的。当发生错误时,领导或负责监督的有经验学生要指出这些错误,并讨论如何避免、更正错误。在实验室会议上,讨论数据收集、删减、分析的错误,能提供了一个绝好的机会,培训学生如何发现和避免错误,而且还可使整个团队产生“头脑风暴”,发现更好的输入数据,避免错误的方法。如前所述,要让每个人都知道承认错误的重要性,不要把这种承认贴上耻辱的标记。毕竟你寻求的是尽可能科学的产品,这需要参与者保持真正开放和坦率的心态。

实验室压力

不管实验室运转得怎么好,也不管主任和实验室成员对问题和批评如何坦诚,紧张和压力必然存在。这种紧张可能来自各种最后期限,如学生的学期论文、考试、会议海报的准备、手稿或基金申请等,有时可能几项同时到期。显然,若能提前做好计划,则会减少这种紧张感。紧张和任务冲突外的其他一些因素也会直接降低实验室的训练质量、研究质量和工作效率。人们在压力下更容易犯错,影响执行计划。所以,虽然一年内不同的时间都可能产生压力,但主任和实验室人员应该直面压力,并尽可能避免和消除产生压力的因素。维持实验室日程安排,列出被试的测试日期和时间、会议和基金的最后期限、考试和论文提交的最后日期,通过这些方法,使整个实验室都有机会讨论工作计划和各种任务的最后期限,以防这些期限总是出现在不适宜的时间或赶在一起。在实验室会议上讨论这些冲突,成员就可在它们发生之前找到避免的策略。

产生紧张的另一原因是人际冲突。一方面,有时候人们不愿意与某人一起工作,这是正常现象;另一方面,很少有人有幸只与亲密朋友一起工作。任何科学家都不可避免地要与自己有冲突的人共事。然而科学研究要继续向前,研究者必须和与自己有冲突的人有效合作。所以,实验室主任和学生一起讨论这些问题就显得尤为重要。主任在这个问题上应该立场坚定,清楚阐明必须遵循的原则和必须达到的效率。这为训练学生成为专业人士提供了机会。一旦学生间发生争吵,主任应该给他们解决矛盾的空间。然而,在某些情况下,为了保证学生的发展和研究工作的质量,主任必须进行某些干涉,这是不可避免的。发生这种情况时,主任要清楚无误地指出讨论和解决问题的准则。显然,学生能从主任和其他专业人员那里获取经验。当发生冲突时,主任必须迅速、专业地与当事人双方私下交流。在公共场合如实验室会议上谈这些问题,会产生不必要的结果,迫使当事双方不妥协,造成实验室内部不和谐。因此,主任应该在训练之初就言明不允许的行为,如对性别、年龄、种族的歧视等。明确行为规则能减少冲突发生的可能性。

实验室的管理层次

新助手进入实验室时，他们总要和“老助手”相处。这就是实验室的“等级”，老助手负责对新生的培训。这是个绝佳良机，可以让老助手学习指导新助手、担负责任以及获得满足感。然而，这有时也会产生紧张感，因为新生不愿意被他们视为同伴的、有经验老助手“指手画脚”。主任在实验室会议上要对老助手担当的训导职责作出说明，把这种训导学生的过程模拟给新生和老助手看，这样就能减少这种紧张状况。实际的分享指导（新来的同学受到主任和高级成员的共同指导）方式就能达到这两个目的，进而区分实验室内不同职责。这能清楚地告诉所有学生，新生将在某种程度上需要有经验的学生指导。

维系一个良好的实验室

一旦经过培训的人员正式加入实验室的工作，标准的实验室规则、坚定的研究传统等因素就会发挥积极作用，保证研究工作更顺利地进行。但是，要对某些因素进行监控，以持续实验室的工作。首先，研究团队应保持明确的程序、职责和目标。模糊性越少就越能更好地开展研究。程序不明确，会使决策变得随意，没有确定的原则来指导。研究目标必须是合理的，在特定的时间内能够完成。即使是长期的项目，也应该有中间的期限。其次，监督必须人性化。学生要能感到他们是整个研究过程中的一部分，他们对实验室的贡献是重要的。主任要在一对一面谈的基础上给予针对个人的反馈，作出鼓励和更正（如果必要的话）。不要低估积极反馈的作用，它们对学生工作质量和长期发展都非常重要。在工作中得到支持和鼓励，他们就会工作得更好，这也有助于我们享受工作的乐趣。同时，实验室主任必须随时指出学生需要更多督导和训练的领域。当学生没完成其职责或某些错误持续发生时，主任应该私下与学生讨论这些问题，明确表达自己的观点，坦诚自己的评价共同制订出改变现状的计划。有时候，学生不能达到实验室要求，或不能处理好工作中的技术问题，主任应向学生指明这些问题。如果有必要，让学生重新参加实验室内部或其他情况下的培训。最不好的事可能就是强迫学生从事某种工作，而他们却缺乏干好这些工作的兴趣、社会能力或专业技术。很显然，主任应预防决策中的这种疏漏。

最后一点，建立一个良好的实验室环境，就要给学生提供表现其主动性和为研究做贡献的机会。在训练学生时，应尽可能早地让学生参与研究项目的设计，鼓励他们阅读有关的文献，尝试提出研究设想。机会成熟时，让学生去参加学术会议，先担当合作者，然后再成为独立工作者。在他们获得进步后，应给予他们上台演讲的机会。这些机

会有助于学生更加投入地其研究训练中，提高他们在学术领域的能力。

建立良好实验室环境的另一个方法就是鼓励学生培养广泛的研究兴趣。学生可能对某些领域的工作或某些项目更感兴趣，实验室主任应关注并鼓励他们在实验室环境下探索这些领域。当实验室项目很多时，这很容易实施。

总 结

本章主要探讨维系一个有效研究团队的要点，包括团队领导的作用、对实验室成员间研究道德规范的沟通、招募研究人员的策略、实验室手册的编制、学生的主动性、在保证研究质量中核查数据的重要性，导致实验室紧张气氛的问题，实验室的层次结构如何发展、影响实验室的研究活动以及定期召开实验室会议的重要性，最后是维系有效研究团队的一些重要问题。虽然以上讨论的一些措施在协调研究团队时非常有效，但这里提供的解决方案并没有囊括全部。而且，这些方法在某些情境下很管用，但换个环境可能就失效了。因而，为维护实验室的有效运转，切实需要实验室主任和团队成员具有灵活性。虽然实验室的声誉一般是建立在实验室主任的工作和威望上，但其事业的最终发展，却依赖于研究团队所有成员的共同努力。

推 荐 阅 读

关于沟通、团队工作以及培训的有相关读物，包括：Tindall（1995），Bormaster & Treat（1994），Johnson（1991），Harris（1993），Wheelan（1990），Larson & LaFasto（1989），Grasha（1987），以及Johnson & Johnson（1987）。

参 考 文 献

- Bormaster, J. S., & Treat, C. (1994). *Building interpersonal relationships through talking, listening, communicating* (2nd ed.). Austin, TX: Pro-Ed.
- Grasha, A. F. (1987). *Practical applications of psychology* (3rd ed.). Boston: Little, Brown.
- Harris, T. E. (1993). *Applied organizational communication: Perspectives, principles, and pragmatics*. Hillsdale, NJ: Lawrence Erlbaum.
- Johnson, D. W. (1991). *Human relations and your career* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Johnson, D. W., & Johnson, F. P. (1987). *Joining together: Group theory and group skills*. Englewood Cliffs, NJ: Prentice Hall.
- Larson, C. E., & LaFasto, F. M. J. (1989). *Teamwork: What must go right/what can go wrong*. Newbury Park, CA: Sage.
- Tindall, J. A. (1995). *Peer programs: An in-depth look at peer helping. Planning, implementation, and administration*. Bristol, PA: Accelerated Development, Inc.
- Wheelan, S. A. (1990). *Facilitating training groups: A guide to leadership and verbal intervention skills*. New York: Praeger.

附录25.1 实验室手册的样本内容

I. 引言

A. 实验室哲学

1. 积极的相互作用
2. 衣着标准

II. 测验程序总览

- A. 邀请被试
- B. 实验前一天
- C. 实验当天

III. 被试招聘

A. 联系

1. 心理学系的本科生
 - a. GEB202
 - b. 其他学生
2. 科特办公室（包括Tandy操作）
 - a. County Clerk的办公室
 - b. Tandy 102 与苹果电脑之间的转换
3. 医院
 - a. 选择被试
 - b. 被试信息
 - c. 邀请被试
 - d. 每天在特别护理托儿所（SCN）所要检查的事项
 - e. 招募从SCN转来的婴儿或被推荐的婴儿
 - f. 与样本被试会谈
4. 重复测试
5. 在前一天打电话预约
6. 对婴儿跟踪测试的程序
 - a. 最初接触
 - b. 在前一夜打电话预约
 - c. 没有电话信件
 - d. 谢绝
 - e. 取消



- f. 更新绿皮书
- g. 给参加研究的小孩发放卡片
- 7. 推荐来的被试
 - a. 婴儿
 - b. 大一点儿的被试
 - c. 医疗资料
 - d. 推荐报告

IV. 行为测试

A. 婴儿发展的Bayley量表

- 1. 呈现的顺序
- 2. 测试中一些小窍门
- 3. 测试1岁儿童时需注意的问题
- 4. 测试2岁儿童时需注意的问题

V. 记录保存

A. 文件

- 1. 医疗记录
- 2. 知情同意书
- 3. 利手表格
- 4. 阻抗表单
- 5. 行为得分表单
- 6. 儿科并发症表单
- 7. 检查列表
- 8. 书面联系方式
- 9. 数据分析记录
- 10. 神经数据表格（针对成年人）
- 11. 听觉感受域值筛选表格（针对成年人）

B. 数据库和记录本（绿皮书）

C. 被试联系方式记录

D. 联系被试

VI. 医院系统

A. 导言

B. ERP设备的操作

VII. 数据的重复检查（十分重要！）

A. 导言



- B. 随机核查
- C. 双份核查
- D. 测试设备变动时的核查

VIII. 通用的实验室测试程序

- A. 提醒电话
- B. 测试日的装备
- C. 为数据数字化架设苹果机
- D. 219-I房间
- E. 电极应用前的准备
- F. 训练被试准备
- G. 测验被试
- H. 实验后的程序

IX. 电极放置——10-20系统 (Jasper, 1958)

X. 保存记录

- A. 联系被试
- B. 被试实验
 - 1. 被试文件
 - 2. 测试信息

XI. EPACS

- A. 描述
- B. 运行EPACS
- C. 峰值分析
- D. ERP地图
- E. 保存CUTT和 BLUM到Cartridge中

XII. 磁带转换

- A. 为微机和主机分析设备准备数据
 - 1. 导言

XIII. 统计分析

- A. EPACS数据例行分析完成后的微机程序和SPSS程序
 - 1. C 语言BASE NORM, MOCK V2.02
 - 2. C 语言AVERAGER 1.39
 - 3. C 语言ERP/CENT Ver7.1
 - 4. C 语言SCHEFFE ANAL Ver.7.1
 - 5. C 语言Z-TEST



6. SPSS Program Package (苹果机 6.1版本)

B. 程序信息

1. C 语言ERP/CENT Ver. 7.1 Program

XIV. 需求

- A. 开具购买发票 (最多100美元)
- B. 购买通知单
- C. 一般储存的购物单
- D. 印刷和复印的货单

XV. 表格

- A. 阻抗列表
- B. 平均核查表
- C. 被试核查表
- D. 婴儿被试的筛选表
- E. 知情同意书
- F. 适龄的NIH被试表格

XVI. 电极放置

- A. 成年人的电极放置
- B. 婴儿的电极放置
- C. 电极放置总结



第 26 章

工作方式的多样性

Elke Brenstein

不同类型的差异

虽然我们不断取得进步，越来越懂得尊重别人性别、种族背景、年龄、宗教信仰和性倾向的差异，但我们还须尊重其他方面的差异，如思维方式、行动、感受、处理工作和对待他人的不同。这种“心理差异”也是人格中的一个重要部分，这种差异并非显而易见，常被其他因素所掩盖。然而，在我们与朋友、家人、同学、同事以及上级的日常交往中，心理差异都起着很重要的作用。以下列情景为例：

在一次会议中，4个人正在做一项计划，旨在改善校园咨询服务。John制作了一个流程图，描述目前提供咨询服务的过程；Kim认为应先坐下来，集思广益地讨论问题，寻找解决方法，这样才合理；June则认为，当务之急是对每个人的工作职责和咨询服务需求作一个详细的任务分析；而Alan则想通过与以前客户和服务人员面谈的方法来获取信息。Alan认为，在对任务或抽象的流程图进行客观分析时，个人的观点会提供更多的信息。John认为，只有设立合理的制度，人们才能各司其职。由于这次讨论由John主持，因此他希望别人都能按照他为会议设定好的议事日程来进行讨论。在某些论题上，John对Kim“不断跑题”感到不耐烦，但Kim则认为自己的论点非常切题，John不应该那么“教条主义”。

旁观者可能会发现，他们每个人的思考都有正确的一面。然而，因为团体中的成员毫不欣赏别人，只是相互激怒。这样的情况很容易导致交流破裂，使很多具有建设性和启发性的意见得不到尊重和采纳。相反，人们在看待事物、构思问题、寻求解决办法上的差异，可能是表现了不同的解决问题的方法，这些方法各有优点和弱点，因此本质上没有什么好坏之分。我们只能根据这些方法是否有利于一个特定目的，来评估它们的价值。

大家可以指责June过于注重事物发展的细节，但同时也可以利用她对细节的判断力和

寻找事实真相的才能。因为June是一个坚持不懈、尽职尽责的人，她做了大量细致的检查工作，汇编了病人和咨询服务的统计资料。大家可受益于Kim对形势敏锐的宏观分析及提出创造性解决方案的能力，而不应去指责她离题。John有时候可能会显得刻板，但他很擅长既客观又批判性地评价一个问题。在这种情况下，由于John不像其他人那样感情用事，所以他可以把那些由服务者人格冲突造成的问题单独分离出来。当然，当Alan评述说，流程图、统计表和创新观均不错，但同样需要考虑其他人员的观点和感情，以便创建一个切实可行的制度，他是对的。因为Alan很有个人魅力，所以，他可获得一些能揭示重要问题的“实用细节”。对一些不情愿提供消息的人来说，这种方法切实可行。

正如上例所示，对一个复杂的合作任务而言，每个人的贡献都很重要。认识到每个人都不同，可使得我们更为宽容和公正；而了解每个人如何不同，可使我们更具建设性地利用个体差异，重视不同观点的价值及其积极作用。本章的目的，是通过回答下列有关心理操作中个体差异的问题，对心理差异的新旧话题进行一个简短而不过于理论化的介绍：

- 它们从什么地方产生？
- 主要的维度是什么？
- 怎么样才能了解自己的偏好？
- 偏好是如何表现出来的？
- 偏好如何发展？

对相关维度和方法作简短总结后，我们将在几个应用领域（学习、工作、与他人交流、时间管理和组织）讨论个体差异的影响。每一个部分包含有例子、实用性建议和练习；其后还有提示部分，告诫大家如何不墨守成规、明智地运用有关个体差异的知识。本章最后有一个带注释的推荐阅读材料目录，该目录为那些想更详细地了解理论，或了解不同领域中实际运用的人提供了参考资料。

心理差异的起源

人们来到这个世上，不仅体形、肤色、高矮胖瘦不同，而且拥有各种不同的“人格”。然而，尽管我们大部分人都可接受这种现象，如一些人身体素质较好，适于跑步、跳高，而另一些人小巧或整体的运动控制能力较好。但是，我们不太容易接受人们在学习、工作、与他人交流能力和偏好的差异，这些差异对个人来说同等重要。

早期的心理学理论（行为主义、精神分析学）过分强调环境影响的重要性，其中华生最有名的论断就是，他能够将任何一个孩子培育成罪犯或者科学家。这反映了行为主义者的一个假设：环境是所有行为变化的原因。但是，随着生物科学的进步，行为的生物学基础不断得到揭示。例如，对分开养育的同卵双生子的研究不断提出强有力的证

据,表明相对于以往的认识而言(参见Tellegen et al., 1988; Neubauer & Neubauer, 1990),行为模式在更大程度上是根据遗传蓝本发展的。因此,每个人身上都有某些生理上和心理上的“天赋”,这些天赋规定了个人发展潜力的范围。通过努力我们会实现更多的东西,但是,由于先天素质的不同,这会影响个人发展的方向和程度。

当然,这个对先天—教养问题的新看法并不意味着遗传决定论,也不会削弱外界影响的作用。因为,“环境”不再被看作是静态的存在,而是一种在与个人交互中动态的变化。在某程度上,“环境”即是我们所感知到的,并且可以根据我们对待社会的方式而作出反应,因此,社会部分是由我们决定的。例如,母亲可能会因孩子以不同的方式对她而会截然不同地对待孩子。按照这方式,人们决定着或者实际上改变着他们所处的环境,而不是单方面地受环境影响。因此,目前有更多的人关注人与环境之间的“适合程度”和“特定的相互作用”。人格测量有助于评估这种“适合程度”。

理论和测量

虽然心理学长期以来一直研究人类个体,但除了人格和能力研究,个体差异一般没有引起过多的注意。相反,在针对具体内容领域的基础心理学和教育的研究中,个体差异通常被认为是“麻烦因素”,成为误差论。研究者通常关注总体的发展模型、阅读理解或创造性解决问题的理论,而不去钻研不同的人会怎样用不同的方式去阅读或解决问题。认知风格的研究确实考察了感知和加工信息上的性质差异,但这些研究通常没有定论。因为认知风格变量测量的是在行为上可能也可能不能明确表达的偏好,它缺乏清晰的预测结果也就不足为奇了。而且,其所使用的维度常常过于宽泛,而不能得出可以解释的结果。最后,缺乏综合的理论框架也使得我们不易评价个别发现的重要性,结果形成恶性循环,个体差异被排除在研究和实践之外,导致理论和研究结果不能解释这些变量可能产生的影响。

大五人格理论

自Costa和McCrae(1992)对英语中的人格形容词进行分析以来,关于人格的个体差异已形成了一个理论框架。他们对人格维度进行了分类。这个工作基于这样一个假设:描述人与人之间不同的所有性质都会表达在语言中(Digman, 1990)。他们识别出5个因素,在众多不同年龄和职业的正常或临床人群中得到了证实。作为此项研究的一部分,逐步发展起来的NEO人格量表,测量了神经质(正常人群的情绪稳定程度)、外向性、开放性、和悦性、负责性。NEO-PI记分以剖面图的形式呈现。尽管NEO-PI不能代表一个专门的人格理论,但是与以其他理论为基础的测量(例如,Myers-Briggs分类指标和兴趣、价值或认知风格量表)的高相关性证实了这个量表的结构效度,也证实它各个维

度的重大价值和普遍性。因为Costa和McCrae的工作是基于纵向数据的，它们也是各个维度在宏观分析水平上相对稳定性的证据。

心理类型

早些时候，在荣格理论的基础上，Isabel Myers开发了一个类似的框架。该理论为后来的Myers-Briggs分类指标（Myers-Briggs Type Indicator, MBTI）奠定了坚实的基础¹。NEO-PI和MBTI共同认为，在某种程度上，功能差异受遗传的影响，因此它表现出某种跨时间和情景的稳定性。这就意味着，人们有某种功能差异（如利手），他们可选择按照或不按照这种功能差异行动。随着年龄的增长和生理、心理的成熟，这些倾向将愈发明显，然而实际的行为可能会变得更加多样化。MBTI一直被广泛地应用于研究和实际工作之中²。在实际应用中，它也是最受欢迎的方法之一。虽然MBTI的有效性已经得到广泛证实，但是科学界并不总乐意接受它（可能是因为它是一个外行人开发的）。由于MBTI的记分方法导致了类型学的出现，所以有些人一直对它持有异议³。在MBTI的记分方法中，根据在外倾性—内倾性、感觉—直觉、思想—情绪、判断—觉察4个两极维度上选择得分，个体被归为十六种类型之一。

尽管人们已经发现NEO-PI和MBTI有高相关，但它们对维度的定义还是不同的。而且，标准形式的MBTI并不评价神经质，因为它主要强调个体差异积极的一面；但MBTI的另一种计算方式提供了一个舒适—不舒适的维度。MBTI或者NEO-PI都是依据不同维度上一系列分数去描述个体，诸如此类的综合计分方法与两个维度分类的计分方法相比，更能提供生态上有效、个体水平上有意义的信息。MBTI和NEO-PI对每个主要维度均可用子量表记分，因此可在较低的分析水平上分析更多的差异。

了解自己的偏好

你知道自己喜欢做什么事情吗？你知道自己擅长什么或者想要擅长什么吗？不管是绘画、下棋、帮助他人，还是处理数字，你都应该找到自己狂热追求的方面，不断去追求、运用和实践，并以此为乐趣，在你自己的学习、职业、私人生活中得到满足。你或许已经知道自己的特殊才能和擅长的方面，也可能并不知道。不过就算知道，你也只能有意识地去运用它们，使其对自己有利。

自我观察

当谈及你工作和学习的行为、动机和兴趣时，你就是最好的专家。想一想你做了什么，怎么做的，为什么这么做，它是否起了作用，你可以把自己的长处和短处、喜好和厌恶列一个表。你可能想通过日记来帮助记忆一些事情，如你在写一篇论文时有怎样的

感受、你碰到什么特别的问题、你是如何一步一步达到目标的、当你与一个同学或同事合作时出现了什么样的人际纠纷。你也应该思考一下为什么你与一些人相处要比与另一些人相处得更好一点，为什么你更喜欢某些老师，或心理学领域中什么特别吸引你。

从其他人那里获得反馈

对你是怎样的一个人，你的朋友、家人或老师可能会有更全面的描述。除了了解你自己外，了解别人如何看待你的行为也很重要。社会是一面镜子，它可以揭示事物的一些盲点，让事物暴露在不同的光线之下。例如，一些人为自己的客观和公正而骄傲，原因是他们能在争议中合理地看待事物，保持“实际的”态度，不牵扯过多的感情。但是，这种态度或许又会被认为是冷淡而缺乏感情的。因此，我们（良好）的用意是一回事，而其他如何理解这用意又是另一回事。

其他资料

有几本关于心理学类型的好书可供查阅（参见本章最后的推荐阅读部分）。部分书中有快速诊断测试，这些测试可能会有助于你更好地了解自己的偏好。当然，与那些由合格专业人员制定的测试相比，这些书中的测试不够精确和合理。如果你想更详细地研究自己的偏好，可以求助专业人员，以获取心理测试题或进行个别咨询。许多大学咨询中心都提供事业和生活方面的咨询服务。

现在，你可以从表26.1开始。它描述的维度以上述MBTI量表为基础（与NEO-PI维度相似）。这些维度根据发表的文献而专门为现在的目的作了改编。为避免使用MBTI或NEO-PI量表可能带来的混淆，这里用了一般性的术语来描述每个维度在不同的上下文中的基本特征。

有一点很重要：术语描述的是一个粗略的框架，为的是从更好理解自己的偏好开始，过渡到更好理解人们之间的差异。描述并不能取代更全面、更彻底的评估。最重要的是，描述并不意味着要给个体“归档”或贴标签。人是非常复杂的生命，有各种各样的兴趣和能力。有某种偏好并不意味着我们不能运用其他不偏好的功能，而仅仅意味这些功能可能需要花费更多的注意和努力。

应 用 范 围

人格差异和爱好、价值观念、技能都影响到我们日常活动的方方面面，从如何去学习、工作和与他人交往，到如何区分任务的轻重缓急和安排时间。本文会对下列的应用领域作更详细的讨论：

- 学习和工作

表26.1 偏好列表

人们的偏好	人们的偏好
人和行为的外部世界	思想和观念的内部世界
大部分人都积极、热情和友善的，喜欢变化和付诸行动，参与班级，喜欢从经验中学习，倾向于出声思考和即兴思考，喜欢通过与他人讨论来阐明自己的思想，喜欢无计划的处理工作，乐意接受分心物，更喜欢面对面的交流。	是安静缄默的，更喜欢一对一的交往，喜欢安静和专心致志，会坐好仔细聆听，喜欢从理论中学习，宁愿在说话前有时间去思考和准备，需要独自一个人才能专心工作，在行动前会认真思考，不喜欢分心，更喜欢书面交流。
具体事实、详细资料和目前的实际	直觉、幻想和未来的可能性
依靠直接经验，关注当前，循序渐进（坚持手头的主题），喜欢精确和实际的信息（只见树木不见森林），因此擅长记忆事实，并且更喜欢首先呈现事实，由实际考虑指引，可能持有这样一个态度：如果不破就不修理它；喜欢依赖可靠和正确的程序来解决问题和应用所学的知识，善于对现有的体制做较小的改动（通过不断改善来创造）。	相信直觉，喜欢思考未来的可能性，思考全面而又有联想性（容易偏离手头的主题），依赖洞察力和想像力（只见森林而不见树木），善于理解事物是如何相关联的，可能由理论考虑引导，喜欢先呈现宏观全景，喜欢用创新的方法解决问题，更喜欢学习新技能，而不是应用它，有时喜欢做彻底的变化和发明做事的新方法（通过发明来创造）。
客观思考和逻辑分析	主观价值观和个人卷入
关注任务，客观而公正地看待一个问题的正反两方面，常基于对事实的逻辑分析做决定，进行论证时可以不带个人因素，可能有点粗暴和挑剔，重视原则，以能力和精于论证而自豪。	关注人，重视人的作用，寻求问题的个人实用性，常基于主观价值观和直觉作决定，进行论证时带有个人信念和热情，重视和谐，以能够促进个人动机或深刻交流信念而自豪。
结构性和明确性	灵活性和选择性
喜欢安排和计划活动，倾向于遵守一个固定的日程安排，关注需要做的事情，用列表来指导行为，不喜欢出乎意料的事，想要坚持到底，由结果指引，重视结果，果断和清晰地陈述他们的目标，去完成可计量的目标，寻求事物的完全性，当某事已经决定或者完成时，感到十分满足。	重视计划中的灵活性，不喜欢狭小的时间路线，可能会背离一个计划好的行动路线，把列表当作新想法的出发点，对于可供选择的办法相当开明，并且能够适应不断变化的需求，注重过程，重视自主权，把他们的观点展示出来公开讨论，经常综合不同的观点，为了质量的改善而工作，把大部分任务看作是进行中的一项。

注释：如何应用这个表格：首先你应通读表格两边的描述。它们是对极限情况的描述，给出了一个两极特征的粗略草图。你可能对双方陈述都会同意，因为你的行为会因不同情景和任务需求而有所不同。不过，你可能会更倾向于其中一方的操作方法。如果你有一个明显的偏好（这个偏好或许与你在实际情景中的行为方式不同），那么你就应该对每对正反方做一个记录；如果你觉得不能确定也没有关系。人格剖面图由四种倾向性的所有组合组成，每种偏好与其他倾向性结合时，其性质都会有一定程度的不同。

- 与他人交往
- 时间管理和组织

在不同领域的心理操作中，人格差异是个体差异的核心。正如Curry（1983）的“洋葱模型”所示，学习行为主要是由一些核心人格维度控制（洋葱的内层），它们转变成信息加工维度（洋葱的中间层），然后通过与环境因素的交互作用而显现出来（洋葱的外层）。研究表明，学习行为和人格维度（如独立开发的工具所测量的）存在大量的共

变 (Geisler-Brenstein & Schmeck, 1996; Geisler-Brenstein, Schmeck, & Hetherington, 1996), 甚至特殊的技能或能力也与人格维度有关。在这种情况下, 很难说是基本能力 (如加德纳的多元智力) 导致相应人格“特质”的发展, 还是与人格有关的偏好引导技能朝某方向发展。然而, 更为可能的是, 在与环境因素持续的交互作用中, 它们相互影响。

学习和工作

正如本章开头的例子所示, 我们可以有许多处理任务的方法。总的说来, 一种做事的方法 (这种方法或许是个人风格的一种表现) 本身并不会比另一种更好。实际上, 最好不要一成不变地只用一种方法, 而是要采取最适合任务和情景具体要求的方法。尽管如此, 我们大部分人确实有一种做事的偏好, 或是因为我们的为人, 或是因为我们已经形成的习惯。因此, 学习或工作风格是指一个人通常在从事学习、工作和与他人交往时的个人偏好。

或许可通过问如下的问题来了解自己的学习偏好: 你通常会同时进行几个项目吗? 你总是喜欢先完成一个任务后再开始做其他事吗? 你通常是先考虑一项计划的细节, 还是先考虑它的宏观前景? 你喜欢从理论上学习事物还是从实践经验中学习? 你重视与同事的和睦相处, 还是很容易超然物外?

写论文的多种方法

尽管好老师会鼓励学生发展个人天赋、兴趣和才能, 但是大部分的学校环境并没有为学生的个性发展提供空间。教育家们经常不自觉地歧视那些不符合“好学生”标准的学生。因此, 与那些更偏爱主动的、富有表现力的、学习方法的学生相比, 那些能静坐几个小时的学生会更符合标准。相对于具体的实际经验, 课程教育更偏重抽象的、符号性的知识; 相对于直觉和个人投入, 课程教育更强调逻辑分析。它常要求学生遵循那些狭义的、规定好的路线, 很少允许学生采用具有对个人更具意义的方法。更糟的是, 学生们有时因为使用不同的方法而受到讥讽甚至处罚。

因此, 教学实践常常反映这样一个信念, 就是有一个最好的方法适用于撰写论文、解数学难题或者完成作业。当然, 大部分人都会同意某种“客观”标准。例如, 一篇好的论文或者一个好的报告必须是有组织、有逻辑论证流程、有实例、以证据来支持观点的。然而, 如以下的案例所示, 采用不同的方式也可以达到这样的效果。

在写论文时, Daryl总是非常有条不紊。他通常是从图书馆中查阅所有相关的信息开始, 复习自己的笔记, 做一个详细的大纲, 然后开始思考所要讨论的问题以及它们的理论和实际意义。

Carrie对某个话题特别感兴趣，他总是在“头脑风暴”会议中写下大体的思想和观点。开始时，论文的结构并不清晰，但当Carrie遵循其直觉和充分发挥想像力时，论文的结构就显现出来了，随后再添加论文的组织 and 细节。

首先运用自己的长处

从上面的例子中可以看到，任何确定的任务甚至是长期目标都可用完全不同的方法来完成，重要的是找到一种最适合你的方法。当处理一个任务时，你应该采用最适合自己的方式或策略，因为它是推动你工作的源动力。通常你已拥有专门的技能，使用这些技能，按自己的方法处理问题，就会使问题解决更容易、更有趣。当然，一旦你已经开始，很重要的一点就是不要忘记其他方面。一个做得很好的计划，应该是覆盖了所有主要成分。换句话说，就是所有下列方面都已考虑到了：

- 理论和观点
- 逻辑分析和批评
- 个人相关性
- 具体例子
- 组织和结构
- 特定事实和细节
- 程序上的细节

与他人交流

当两个人看起来似乎不能相互理解时，问题并不一定出在他们说的内容上，而常常是出在对要说的内容的表达上。这不仅涉及表达信息的语气，而且涉及更多的人与人之间的差异，这些差异或许是“交流问题”的根源。

George在学生处工作，由于和上级存在“交流问题”而感到非常沮丧。他有时到老板的办公室去，“侃”一些创新性想法。问题是，他的议论如泥牛入海，老板似乎从没认真听过他在说什么。同时，George与同事Jarret有过许多私下的讨论，当他们谈论和思考未来发展时，他的想法变得越来越具体。为什么老板不能够理解他的想法的价值呢？在与老板参加一个MBTI的研讨会后，George弄明白了老板是一个更喜欢书面交流的人，喜欢有时间把事情再思考一遍，做事非常有组织、有条理，强调逻辑性和细节。从此以后，George把想法写下来，确保这些想法都经过很好的思考。同时，对不同的可能性，做出合乎逻辑的概述，并且提供有助于鉴证其建议可行性的必要事实。立刻，老板对George的想法充满了热情，并且召开会议讨论如何实施这些想法。

诸如以上案例中的“交流问题”可能严重地影响一个团队中的工作气氛，对研究结

果造成负面影响。在这个例子中，George了解了老板为什么对他热心提出的想法不予理睬，找到了适合的解决方法，避免了与老板的争吵。因此，我们不仅要了解自己的工作方式，也要认识到别人的工作方式，交流的质量就会有很大的改善。尽管这种认识本身不能解决问题，但它给双方提供了一个讨论差异和争论的基础，并可以采取避免这些差异和争论。

在团队作业中，能意识到并且能正确评价差异是极其重要的，这样可以使每个团队成员都能够贡献自己的看法和技能。如上所述，当独立工作时，我们尝试做所有的工作；当作为一个团队工作时，我们可以把任务分开，让每个成员处理问题不同的方面：

- 清点存货
 1. 问题是什么？
 2. 目前的事实是什么？
- 形成对未来的远见
 1. 事实意味着什么？
 2. 解决问题的不同可能性是什么？
- 用批评的眼光评估不同的可供选择的方法
 1. 每个备选方案的正反两方面是什么？
 2. 每个备选方案的逻辑结果是什么？
- 考虑人的影响
 1. 不同的人如何不同地评价每个可供选择的方法？
 2. 对每个结果人们如何反应？

人们建设性地、和谐地运用他们的才能所产生的协同作用，可极大地改善研究过程，提高研究成果的质量。意识到看待事物有不同的方法，这些方法本身各有其价值，对这种意识的敏感性可能是避免“交流问题”的第一步。如果这种意识能一直敞开交流的渠道，我们就获益匪浅。

第二步是了解别人是怎样看问题的。了解潜在的人格差异为我们提供了一个框架，让我们可以建设性地讨论这些差异（然而，应注意，试图诊断别人是困难的，而且有贴标签的危险；参见本章结尾的提示部分）。为达到这样的效果，可让每个参加者都写下一些东西，然后再讨论各自不同的观点。如果某个“问题”已经偏离问题实质而进入个人范畴（“你总是离题，你总是对一些细节吹毛求疵”），那么听听别人是如何评价他们的所作所为，可能有助于了解产生差异的原因。在这一点上，以下的练习可能对你有帮助：

你有时是正确的，我也并不完美

1. 指出你所欣赏的别人身上的素质，试着考虑不同的方面。

2. 花一些时间，思考你处理问题、与他人交际的方式是否有问题。
3. 与你的搭档分享自己的成果。

通过提高聆听技巧和从别人的视角看问题，我们可以获得很大的收获。了解别人的背景也可以更容易“求同存异”。如果遵循这个练习过程的讨论仍不产生效果，那么参加一个由专业人士主持的研讨会或接受高技巧调解人的帮助也许就十分必要了。

使用一种特殊的做事方法（如宏观考虑）并不意味着你就不能考虑细节。这仅仅意味着对你来说，运用自己偏好的方式更容易，而采用不偏好的方式可能需要更多的努力，或许还需要一些“窍门”。例如，把每件事情放在一个日常计划中来思考细节问题。解决“问题”的第一步是了解这个问题到底是什么。随着时间和练习的增加，你就可以得到一个更为和谐的工作方法。

时间的管理和组织

管理时间非常容易！你要做的就是弄清楚需要做什么，按照事情的先后顺序对每件任务赋予优先编码，然后只需去做就可以了。以逻辑顺序来核对项目，事情做得又快又有效率。有关时间管理书籍的作者就是这样告诉我们的。然而，如果真这么容易的话，那就不会有这么多关于这话题的书了。

有趣的是，有些人认为很容易用这种方式操作，并且想给别人展示如何做，包括写书的人。他们不承认别人有根本不同的处理方法，并且这些方法可能对他们同样适用（甚至更好）。并且，他们很少关注有效率的、以任务为中心、以结果为导向的运行机制方面，更不会在研究方法、程序、检查任务目标的合理性方面考虑为人们的感情留下空间。

用上述的逻辑顺序分析你在哪里失败和为什么会失败，可能会帮助你处理好时间管理问题。你应做一个个人化的行动计划来开始工作，而不是因为不能完成看似合理的要求而指责自己。尽管有些人很轻松就能组织工作、坚持时间安排，但大部分人都会因为这样或那样的原因，推迟或拖延一个任务。正如你可能已猜到，人们常有不同的原因。下面以某人为例：

人们很容易把Julie描述成一个外向的人。当Julie与别人一起工作时，她积极参与且很有干劲，但总是把诸如读书、学习的“安静型”任务往后拖延。因为她喜欢独立、即时思考，所以她会避免提前计划。她的室友Sylvia是一个较内向的人，会花更多时间去思考想做的事，并且经常推迟实施自己的想法。

Howard是一个爱动手的人。在学习时，他常对那些太理论化的阅读感到厌烦，但他不介意做那些诸如打印参考资料的常规工作。在工作团队中，他常常因为在细节上花太多的时间和用一种确定方法拼命工作而受到指责。相反，他的搭档Jose在写论文时，对一

一个新题目非常兴奋，喜欢和朋友海阔天空地讨论。因而，他会拖延那些与实施想法有关的日常工作。由于他讨厌琐碎的工作，因此他常常请朋友来给他的论文校稿、帮他整理参考资料。

Alice非常善于领会需要做什么、以什么顺序做。由于她是一个很有逻辑、很理智的人，因此不容易分心。她的朋友常提醒她放松一点，不要总是这么严肃，可是她却很少给自己留出玩的时间。Alice的朋友Ellen尽管也把要做的事情列成表，但是却经常转移目标。一天，Ellen正在复习功课准备考试，一个朋友因个人问题打电话。Ellen和朋友的长谈占据了大部分复习时间，Ellen在很短的时间内轻松地对事情作了重新安排。在这种情况下，Ellen考试可能得不到A，但她认为在朋友需要时及时给予帮助更为重要。

Tammy是以擅长组织而闻名，她做事非常有计划并且很果断。完成任务时，她每次只做需要做的部分，在开始做另一部分之前一定完成先前的任务。这种方法很适合她，然而，同事却抱怨她没有考虑别的问题，做事太草率。并且，当情况突然发生变化时，她很难适应新的环境。Tammy与朋友Sara正好相反。Sara总是同时操作大量不同的计划。她喜欢这种做事的方式，各种各样的事情都很感兴趣。一般来说，她偏重于宏观控制，喜欢考虑事情是如何相互关联的，因而能对论点做出详尽的研究和深刻的描述，并受到导师的称赞。由于她对许多事情感兴趣，所以很难区分工作的优先次序，并彻底处理好一个问题。

正如这些例子所示，实际上并不存在一种最好的方法去组织时间和个人资源。重要的是，每一种操作方法都有正反两方面需要考虑。

回到时间管理专家提供的诀窍上，很明显，每人都需找到自己的规律，并且符合每个人做事的价值观及方法。

与时间管理一样，对事物的组织似乎也存在着公认的标准（如当你需要某物时你能找到它）。但是，重要的是要认识到，对事物也有不同的组织方法。例如，你可能认识这样一个教授，在他办公室的桌子和地板上有成堆的书和文献资料。然而，当你要找两个学期前写的一篇论文时，他可以准确地从一堆书中抽出你的论文。研究证明，有些人不能很快在许多归档的文件中找到一个文件，只是因为他们的不是按那种方式组织知识。当他们把东西归档后，常常记不住东西是被归档在什么标签下面。下面的推荐阅读部分列了一些书目，专门讨论了那些“擅于组织的人”所存在的特殊组织需要。

平
常
学
习
PDG

成长和变化

拥戴类型

Myers已经杜撰了“拥戴类型”这个术语来强调为个性自豪的重要性。在成长为一个成熟之人的道路上，第一步就是注重和开发自己的长处。如果允许我们遵循自己的偏好，并实践伴随这些偏好的技能，我们就可以学会最大限度地发挥自己的长处并有信心接受不喜欢的活动的挑战；如果不让我们在个人或职业中“成为我们自己”（心理学类型理论把此称为“类型证伪”），那么这种能量就会消减，自信心可能受到打击。例如，一个外向型的孩子，常常因为太闹受到内向型的父母责骂，这个孩子在成长过程中可能会觉得自己身上有“某些不对的地方”。同样，在一个结构严谨、严格遵守课程、不会对学生的创造兴趣和能力提供出路的学校环境中，一个具有灵活性、倾向于直觉和联想思维的学生就会四处碰壁。

生活选择

随着年龄的增长，我们通常会处于一个更好的位置，能有意识地选择与自己偏好相符的环境。例如，研究表明，具有某种人格、兴趣和价值观剖面图的人更易被某种职业吸引。当然，这只表明了谁会被吸引到某种职业中去。实际上，具有不同人格剖面图的人在一个特定的职业中都有可能成功。由于诸如心理学或医学之类的许多行业其内部变化多端，在兴趣和能力方面的要求弹性很大，在一个职业领域中发现适合自己的环境，了解人格偏好将很有帮助。因此，有些人更有可能选择某些诸如咨询、临床、实验心理学的专业，或这些专业中的某个专项，如理论工作或与客户打交道（参见Leong & Geisler-Brenstein, 1991）。

与不同想法的人相处或因不适合的环境而产生的问题，我们不应过于关注，应该建设性地运用这些差异。一旦遇到较少偏好的活动，我们就要有意识地补偿。例如：

有一个教师，在小群体中表现很出色，但在300个学生的大群体面前却很难表现出机智和幽默。那么他就需要额外花些时间，想些奇闻趣事，并把它结合到一个枯燥乏味的演讲中，也可以拿出一个录像，或者让班里的学生准备一些东西与大家分享。

只会用打骂来教育子女的内向型父母要保证他们外向型的孩子有足够的时间与朋友一起玩耍。

在某种合作关系中，一个可能会仔细安排整个假期每一分钟的伙伴应留下几天时间，不做安排以适应别人临时做这件事的需求。

倾向于理论化和抽象思考的人在作报告时，可以努力想一些实际的例子。

工作和娱乐

经验表明，当我们运用自己偏好的方式工作时，工作效果最好。正如上面所建议的，如果我们不喜欢当前正在从事的工作，就需要坐下来研究一下，究竟喜欢做和擅长做的事在多大程度上类似于我们正在实际从事的工作。对工作不满的另一个原因可能是我们还不能在工作和娱乐之间找到恰当的平衡点。当我们寓工作于乐的时候，就是我们乐于工作的最好迹象。

Csikszentmihalyi (1990) 杜撰了“涌流”这个术语，来表示人们在活动中可能遇到的一种特殊经验。一个外科医生可能在手术中体验过涌流，一个统计学家在做数据分析时可能会十分投入，一些人在写论文、授课、爬山、读书或者给花园除草时可能会完全忘记时间。涌流体验是一个高度的专注与愉快感觉相结合的综合体。当一个人的工作不受打扰时，当任务可以有多种多样的结果（包含使注意力集中于身体的重复运动）时，涌流经验就很可能发生。任务需要难度适中，这样才易于掌握。任务太容易就没有挑战性，容易产生厌倦；任务太难可能会让人感到挫折。

在娱乐和从事业余爱好时，运用较不偏好的方式也可以发生涌流。例如，那些在日常生活中更依赖个人价值而不是逻辑分析的人，可能会在业余玩策略游戏（如国际象棋）中倍感快乐；一个在需要注重细节及精确性的高度结构化行业中工作的人，可以在水彩绘画中找到快乐，找到创造力的出口。总之，在工作和娱乐中找到平衡，在偏好和不太偏好中找到平衡，真的非常重要。在学习、工作、与别人交流中，成为一个“多功能”的人，可以使我们适应各种各样的需求。

提 示

下列几点要记住：

- 我们可以根据任务或情景的需要而以不同的方式行动。然而，我们大部分人会偏好某种行为模式（参见表26.1）。
- 一个人格“类型”或“剖面”来自于几种不同偏好的结合。因此，把一个人描绘成内向或者外向就过于简单化。
- 尽管有证据表明，某些特征具有跨时间的稳定性，这些特征决不应被用作取舍某事的理由。相反，我们应随着成长和进步而应尝试开发较少偏好的功能。
- 类别不应成为刻板看法或印象的理由。就算是那些具有大部分主要特征（某一个人格类型）的人，在许多方面仍然有所不同。
- 工作方式不存在对或错；然而，在特定的场合中，特定的方法可能会更适用。
- 你应该为你拥有的力量而自豪，并应最大限度地利用它们。

- 你应该意识到自己那些较少偏好的能力，尝试去增强它们，找到弥补弱点的方法。
- 尽管我们不能轻易改变自己一些习惯性偏好，但我们确实可以选择，全面开发各种才能。

推 荐 阅 读

对那些想更多了解“大五”人格的人，NEO-PI手册（1994）是很好的入门。它解释了其构造和记分程序，并且有在不同领域中应用的例子。对于那些想更多了解该主题的人，此书同样也提供了很多参考资料，比如：想了解人格的毕生发展，可参见McCrae和Costa's《成人期的人格》（*Personality in Adulthood*）（1990）；那些对遗传学和人格感兴趣的人可参考Neubauer和Neubauer的《天性的个性特征：新人格遗传学》（*Nature's Thumbprint: The New Genetics of Personality*, 1990），该书介绍了大量分开抚养的同卵双生子的研究。

还有几本关于心理学类型理论的书。对研究心理学类型感兴趣的人而言，Myers和Myers的《天赋差异》（*Gifts Differing*, 1980），一书已成经典著作。此书介绍了MBTI中所隐含的理论，并且讲述了各种领域中不同类型的分布。最近，Hirsh和Kummerow的《生活类型》（*Life Types*, 1989）使更多公众可以理解心理学类型的概念。他们的书提供了一个小小的诊断测验以及类型在不同生活领域中表现出来的例子。读者可在Myers和McCaulley的《MBTI手册》（*MBTI Manual*, 1985）中和Thorne和Cough的《类型的描述：关于MBTI的一个研究纲要》（*Portraits of Type: An MBTI Research Compendium*, 1991）中，读到对MBTI的介绍以及它与别的测量方法的相关性。最后应指出，一些作者在探讨气质差异（依据类型偏好功能的综合）时，对心理类型持不同的观点，例如Keirsey和Bates的《请理解我》（*Please Understand me*, 1984）。对MBTI的批评可参见Pittinger（1993）、McCrae和Costa（1989）的著作。

现有几本书谈及怎样在不同的情形中运用类型理论和MBTI。例如，Provost和Anchors'（1987）的《Myers-Briggs的类型指标在高等教育中的应用》（*Applications of the Myers-Briggs Type Indicator*）讨论了类型知识在指导学生、决定个人学习风格等不同领域中的应用。Provost（1990）的《工作、游戏和类型》（*Work, Play, and Type*）考察了工作和游戏之间微妙的对比，并借助许多练习和例子，展示了类型知识如何有助于在这两方面都取得令人满意的效果。在《你到底是谁》（*Do What You Are*）一书中，Tieger和Barron-Tieger（1992）考察了心理学类型对职业选择、找到适当位置的涵义。这本书提供了一个小测验，并且包含许多实用的小窍门和建议。

除了许多关于时间管理和组织的传统书籍，还有一些书是为“其他人”写作的。尽管下列书不是从心理学类型的观点出发，但它们探讨了用不同方法思考和不同组织的人

所具有的特殊需要。例如, Lehmkuhl和Lamping的《为了创造性的人而组织》(Organizing for the Creative Person, 1993)区分了“Arbies”(右脑创造性、发散思维型)和“Elbies”(左脑、理智的、复合思维型)的组织风格。同样, McGee-Cooper(1992)告诉我们,“你不必带着工作的疲惫回家”,为如何在做好工作的同时留下创造力和游戏的空间提供了实用的建议。

注 释

1. 参见Myers和Myers(1980)的导言部分。
2. 参见Myers和McCaulley(1985)与Thome和Cough(1991)的研究结果概述。
3. 对于有关类型学的适当性、理论性争论,参见McCrae and Costa(1989)。

参 考 文 献

- Costa, P. T., & McCrae, R. R. (1992). *The Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.
- Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper & Row.
- Curry, L. (1983). *An organisation of learning styles theory and constructs*. ERIC Document, 235 185.
- Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, 41, 417-440.
- Geisler-Brenstein, E., & Schmeck, R. R. (1996). The Revised Inventory of Learning Processes: A multifaceted perspective on individual differences in learning. In M. Birenbaum & F. J. R. C. Dochy (Eds.), *Alternatives in assessment of achievements, learning processes and prior knowledge: A European perspective* (pp. 283-318). Boston: Kluwer.
- Geisler-Brenstein, E., Schmeck, R. R., & Hetherington, J. (1996). An individual difference perspective on student diversity. *Higher Education*, 31, 73-96.
- Hirsh, S., & Kummerow, J. (1989). *Life types*. New York: Warner.
- Keirsey, D., & Bates, M. (1984). *Please understand me: Character and temperament types*. Del Mar, CA: Prometheus Nemesis.
- Lehmkuhl, D., & Lamping, D. (1993). *Organizing for the creative person*. New York: Crown.
- Leong, F. T. L., & Geisler-Brenstein, E. (1991). Assessment of career specialty interests in business and medicine. *Career Planning and Adult Development Journal*, 7, 37-44.
- McCrae, R. R., & Costa, P. T. (1989). Reinterpreting the Myers-Briggs Type Indicator from the perspective of the five-factor model of personality. *Journal of Personality*, 57, 17-40.
- McCrae, R. R., & Costa, P. T., Jr. (1990). *Personality in adulthood*. New York London: Guilford.
- McGee-Cooper, A. (1992). *You don't have to go home from work exhausted!* New York: Bantam.
- Myers, I. B., & McCaulley, M. H. (1985). *Manual for the Myers-Briggs Type Indicator: A guide to development and use of the MBTI*. Palo Alto, CA: Consulting Psychologists Press.
- Myers, I. B., & Myers, P. B. (1980). *Gifts differing*. Palo Alto, CA: Consulting Psychologists Press.
- Neubauer, P. B., & Neubauer, A. (1990). *Nature's thumbprint: The new genetics of personality*. New York: Addison-Wesley.
- Pittinger, D. J. (1993). The utility of the Meyers-Briggs Type Indicator. *Review of Educational Research*, 63(4), 467-488.
- Provost, J. A. (1990). *Work, play, and type*. Palo Alto, CA: Consulting Psychologists Press.
- Provost, J. A., & Anchors, S. (1987). *Applications of the Myers-Briggs Type Indicator in higher education*. Palo Alto, CA: Consulting Psychologists Press.
- Tellegen, A., Lykken, D. T., Bouchard, T. J., Wilcox, K. J., Segal, N. L., & Rich, S. (1988). Personality similarity in twins reared apart and together. *Journal of Personality and Social Psychology*, 54, 1031-1039.
- Thorne, A., & Gough, H. (1991). *Portraits of type: An MBTI research compendium*. Palo Alto, CA: Consulting Psychologists Press.
- Tieger, P. T., & Barron-Tieger, B. (1992). *Do what you are*. Boston: Little, Brown.

第 27 章

申请研究基金

John G. Borkowski

获得研究基金并不难，但要有一个好的想法，制订一个行动计划，愿意花大力气，然后找几个合适的基金会，做好心理准备，经受住一开始的挫折，坚持下去，直到你得到资助为止。

虽然获得基金很容易，但是大部分年轻的学者却申请不到，这主要是因为他们害怕失败，而不是因为他们思想浅薄。那些争取基金支持的人之所以最终失败，是因为他们经历最初的失败后便放弃了。因此在基金的争夺中没有获胜的主要原因在于个人动机，而非智力因素。

本章提供了一些申请研究基金的窍门以及不达目的誓不罢休的动机。首先本章将阐述申请研究基金的基本原理，然后提出一些关于如何书写你的第一份基金申请或改善你的再次申请的实用性建议。

为什么要写研究基金申请

我不相信美国在未来10年内会出现这样的一位青年博士：他立志在学术界或职业生涯中取得成功，而他的工作却无需写作和赢得研究基金就可取得显著进展。这种情况在学术界显而易见：职称委员会总是对有一系列基金申请成功记录的助教印象深刻。有些人认为，成功的申请基金记录与坚实的研究成果记录同等重要。这通常对学院院长来说至关重要，因为在某种程度上，申请基金的一些间接耗费所带来的回报会促进财政周转，当然，对负责研究工作的副校长来说，这也同样重要，因为你的基金能使他监管的研究领域扩大了疆界。

更重要的是，研究基金使你能够实现一个长期的研究计划，这比没有基金支持的计划更持久、更系统。你会发现你有更多的研究时间；你的研究就会随着投入的增长而表现出更高的水平。例如，你可能连续3年每学期少教一门课。当然，你也可以雇用更多的

助手。在你的基金预算中，增加招收研究生和本科生可以注入新的活力，这不仅有助于实施研究计划，更重要的是有助于修改、重新考虑和从事新的研究方向。你对自己的职业生涯也会有更多的、更灵活的控制权，因为你不再需要去向你的系主任申请计算机、供应品和旅费了。但最重要的是，你将更有信心进入在你的领域，与“无形的学者群”并驾齐驱。这种进入会为你构建或加强学术关系网，增强合作的可能性，同时你的自我效能感也会增强。所有这些都来自于你的第一个研究基金所带来的良性效应，它们将扩大你的短期研究项目，同时对你长期的学术生涯也大有裨益。

申请研究基金的最后一个原因与你想法的质量和范围有关。在当代心理学的大部分领域，简单的研究课题所剩无几。更可能的是，你的想法在理论上是复杂的、多面的、以过程为导向或者在本质上是纵向的。如果确实如此，那研究将是一项艰巨的任务，需要大量的人力和物力，需要资金支持才能顺利完成。虽然在心理学的一些领域，不用外部资金仍有可能维持一个持久的研究活动，但毫无疑问，基金的支持会促进你的项目的活力和竞争力。

很多年轻的、有抱负的临床医师和咨询师会认为，关于基金申请的这一章与他们未来的职业生涯无关。我不这么认为，很大程度上是因为我看到，很多专业人士获得了为示范、计划和评估项目申请研究基金所需的技能后，在心理健康机构中的管理层青云直上。甚至对每一个在家里、学校、门诊中治疗的病人，如何防止意外发生的新想法，都需要一份包含可执行性、准确性、结果范围这几个要素的详细文件。

对所有的职业心理学家来说，具备写基金申请的技巧（和意愿）都是一大优势。我总是推荐在咨询师和临床医师训练中表现出色的学生去听统计和研究方法的课程（不是像通常的那样截取其中的一小部分），学习基金申请的技巧。预先得到警告就是预先得到装备：回到大学课堂重新捡起“消逝的”计算机、统计和研究方法课程往往为时已晚，然而对很多人来说，这种知识和技能以及对基金申请过程的了解会非常有用，对于职业发展来说可能是必需的。

你寻找，你就会有所发现

想出好的创意

如果我能告诉你如何想出一个有趣且具有理论重要性的想法，那我可能将是一个百万富翁，或至少在自己的研究领域中已有很多好的创意。问题是，好的研究思想来自于某个领域中的现有知识、相近领域中的知识以及技术技能这三者之间的一些未知组合——所有这些都通过灵感、想像和运气融合在一起。只有亲身体验过才会知道，形成一个具体且重要的研究思想所需的因素是如何汇集到一个科学家的头脑之中的。然而，我的确知道，想要拓展你的研究领域的真诚志愿、为了某种目的（而非自己的）探索领域

前沿的意愿、坚持不懈的决心、认为科学游戏很有趣的看法，将使你处于准备和灵感的有利地位；再结合运气，就可产生一个新颖且重要的研究思想。

请记住，原始的思想需要不断精炼和反复琢磨才会成为研究假设。通常的情况是，年轻的学者在他们的基金申请中勾勒出初步的想法，而不是由预实验数据支持的、在理论上精炼而有效的思想。不要害怕让你的思想处于混乱状态，要与他人讨论，有时候你不得不承认，你最初认为是一个很好的创意，其实并不适合研究或根本没有意义。不断努力，过一段时间后，你就会得到最好的想法。然后用热情和决心，在一系列检验假设的研究中付诸实践。

听众和思想

不管基金是用来检验一项理论还是进行一项干涉项目，如果脑海中没有一个确定的基金申请机构，你几乎得不到一项研究基金。你不仅应将你的思想与基金申请机构相配合，而且还要与潜在的评阅者相配合。也就是说，你应该确定并保证你的提案属于一个锁定的基金会或出资单位所要赞助的范围。如果可能，试着找出那个将审阅并最终评定你基金申请的人。写申请时要针对特定的听众。

如果你的计划属于一家大型的联邦机构的赞助范围，那么，国家科学基金（NSF）、国家健康研究院（NIH）、国家精神健康研究院（NIMH）的现有的手册和项目官员目录会为你提供必需的信息。如果你要回应邀请——称之为“邀请研究计划”——那么联邦注册局可能会提供详细的信息；你所在研究所的基金办公室可能已预订了联邦注册局的服务。你们的基金会官员将提供与你的工程相关的私人基金机构的信息。

如果你计划申请私人基金，应该采取如下步骤：（1）通过阅读此基金会最近出版的关于基金决策的年报，确保他们的赞助重点与你的研究兴趣相一致；（2）写一个简短的2~3页的摘要，概括介绍与基金会目标相关的计划；（3）最重要的是试着“踏进门槛”。也就是说，请一个朋友或同行，在你的申请送达之前，先通过电话向相关的基金负责官员介绍一下你的研究，吊起基金会的胃口。不经申请、不经预先的人际接触，便得到基金支持的情况十分罕见。如果你与基金会没有直接联系，最好进行一次拜访，这样官员们才能逐渐了解你，并感受到你對自己研究想法的热忱。

这里讨论的、至关重要的一点是使你意识到，你需要以恰当的方式来表达你的想法，以吸引评阅人的方式来塑造想法的必要性。你需要毫无偏离地跟随机构的指南（任何部分都不要超过页数限制，附录中材料不能过多而增加评阅人的负担）。记住要根据基金会的而不是你自己的要求和特性来表明想法。

形成研究计划

一般而言，研究计划在你头脑中是日渐成熟。在申请早期先试着列出计划总体提

纲，然后一遍又一遍地考虑这些想法，将其不断完善。在尝试过程中，要灵活地多方位考虑你研究的定位，可以从狭义层次上（如检验一个特定的理论），也可以从长期应用和实用价值方面。既然联邦机构及过去仅资助基础研究的私人基金越来越重视有社会性导向的研究，那么也许有一天，如果没有足够的对理论导向研究长期应用意义的讨论，就不太可能得到资助。

你的研究计划有两个目标——理论重要性和潜在的意义，前者应受到更多关注。很简单，必须使评阅人确信，你的研究将提升某个特定范围内的现有知识层次。要做到这一点，至少有3种方式：（1）展示合理的、连贯的论据来支持你的理论，清晰表明其与先前观点的不同之处；（2）十分仔细地讨论你的目标、研究问题和一般假设，清晰地表明它们之间的关系；（3）呈现你的背景文献，特别是你在此领域内的最新工作。第3章和第6章在这方面提供了更多的指引。

最后一步，表明你的研究能力。这是基本的一步。因为你必须展示出令人信服的学术水平（如最近的、相关的出版物），或至少用有力的预实验数据来表明你对实验的理解。在进行大规模的临床示范或干预治疗时，你的学术证明可以是进行小型干预治疗的预实验或一个个案研究表明，整个计划的确可行，具备改变变异或延迟行为的极大潜力。在你写完这部分后，一定要提出特定的假说（提出主要的就行了），也要给出你对研究或者干预的主要结果的预期。

研究细节

研究基金申请的第二部分是研究计划。它在很多方面都与杂志文章的方法部分相似，由下面几部分组成：被试、设计、材料、步骤和准备做的统计分析。如果计划包含多个实验，你需要对每个研究都表达这些信息，除非那些重复、累赘的信息。

被试

基金申请者应尽可能详细地描述被试的数量和最重要的特点。例如，如果被试是大学学生，除了被试的数量、性别、年级和能力或专业成绩信息之外，并不需要更多的信息。另一方面，如果你要研究智障儿童，那么研究应该包括被试的数目、性别、年龄、智商、障碍类型、相关的运动障碍或视觉障碍、送进专门机构治疗的时间、阅读能力、药物治疗史和任何其他重要的特点。

记住要详细描述被试特点，以使评阅人能够从你的研究样本推广到合适的人群上去。通常，获得被试的方法——义务的或付费的，智力正常的还是有临床障碍的人群——会影响实验结果，因此这些信息都应在计划中出现。还有，被试是否受过针对任务的强化训练、以前是否参加过类似研究、是否对实验一无所知，这些都应写进被试部分。第9章关于申请使用人类被试的内容勾画了一系列所需描述的细节。

设计

如果你的研究包括一大堆自变量，每个被试有好几个水平，那么有时候把完整的变量关系告诉评阅人是明智之举。不仅要对变量进行说明，而且还要提到变量属于被试内还是被试间设计。如果已在其他地方详尽描述过自变量，或自变量的数目非常少，那么通常会省略设计部分。然而，通常在一项基金申请中，你要进行新异及复杂的设计，那就需要把这部分作为申请材料的一个重点来写。根据统计效力可计算出检测到合理结果所需的被试数量，因此也应增加这部分的分量。

器材和材料

器材部分应描述器材的型号（品牌名称），如非商业性供货，还应描述它的特点和物理特性。器材部分应该告知设备做什么用的，而不是它们是怎么组装在一起的。应包括主要的应用范围和功能操作，而不是仪器的照片或图表。如果材料很复杂，如在对30个项目的自由回忆中所表现出的类别相关词的数目和多少，那么它们的特性就应该另作介绍，作为一个专门的“材料”章节。

步骤

将读者放在被试的位置上来考虑，是写好步骤部分的最好方法。从被试开始实验的第一点到结束实验的最后一点，一步步地交待。流程图对组织事件的顺序非常有帮助。步骤部分应该准确描述将要展示给被试什么和要求他们做什么。一般而言，最好注重描述被试的行为而不是主试的行为。例如，在一个测量反应时的研究中，假定实验者在每一项目测验后需要重新定。对重复实验时的重要细节，如事件的顺序、持续时间、对被试的指导语（解释指导语而不是逐字汇报，除非指导语是需要操纵的变量之一）、测量的类型和控制的事件，都应在步骤部分报告。请记住，只汇报那些进行重复实验时所必需的细节，避免“步骤”这一节及项目的各研究部分之间的重复描述。

分析和预测

记住这部分是对你期待得到的结果的言语阐述，这些结果要得到数据统计分析的支持。目前流行使用统计技术，但要将分析数据所需的复杂程度降至最低。换句话说，除非是你的研究设计和假设需要，否则不要太复杂。要关注条件间平均值的比较，同时也要注意对个体差异的分析（如果理论上重要的话）。个体差异分析常常可为主要假说提供支持，而主要假说则集中在组间均值的比较上。因为大多评阅组都包括统计专家，所以，在这部分上花些工夫，并由本单位的统计学家过目是很重要的。

附加要点

申请的最后一部分应包含使用人类被试的声明（如出现在附录中的知情同意书、被试中要含有适当比例的妇女和少数民族，后者是某些机构如NIH的要求），相关内容参见第9章。

“研究重要性”应在申请书的最后一段。它给你最后的机会，表明你的工作在此领域可能产生潜在的、长时间的影响，以及为什么资助你的研究会造福科学和社会。当然，自传部分是你申请书的最后部分。你应准确完成这部分，并遵从现在的美国心理学会APA格式（APA，1994；参见第21章，专门介绍APA格式）。一个不符合格式要求的自传会被认为是研究者“马虎草率”的表现。

基金申请最重要的部分：摘要

基金申请的起始部分是摘要。摘要应总结操纵的变量、主要任务、重要的步骤特性、理论相关性和潜在重要性。我故意将第一部分放到最后来讨论，目的在于强调，在写作过程中，摘要写作是整个工作的最后一步，也是最重要的一步。相比较而言，放到其他部分完成之后再写也最容易。

简洁、精确和凸显理论重要性是一个成功摘要的主要特点。在一个好的摘要所包括的所有重要信息中，主要假设以及如何验证它是最重要的。短短250个字以内，达到这一目标而同时又包含设计的特性、主要变量和重要性，并非一件易事。最后不要忽略，一个好的摘要，就像文章的题目一样，可引导评阅人去更为深入和严肃地审阅你的申请。这就需要你精心选择词句和短语来组建摘要。它应是申请最有趣、构思最细致的部分。记住，有些评阅人可能只详细阅读你的摘要，不会直接阅读你的申请；不要错过让他们对你研究的重要性和项目主题留下深刻印象的机会。

准备预算

很多年轻的研究者不恰当地花费大量的时间在预算准备上。预算应该充足但不能过分：你的薪水（包括工作年度和暑假）、雇用者的薪水、装备、供应、旅行和被试费是预算中直接花费的典型项目；研究者也经常考虑到通货膨胀，并为将来的年份增加资金（如4%）。间接的花费（由你的大学和机构决定）加上直接花费，就是本年申请基金的总额。间接花费常由基金申请官员协商决定，但不会超过10%。对任何非正常的花费项目（如为采集数据而寻求专业帮助）都需全面分析其合理性。如存在疑问，要另外花些时间来确认，而不能给你的评阅人留下任何疑问。

获得资助的小窍门

怎样开始

基金申请准备应是你日常职业生活中的一部分。在你的桌子上准备一个笔记本，随时记下你的新思想以备后用，精炼并扩展旧的条目。开始时谦虚一点，将你的第一个申请寄往学校内部或本地的机构。简要读一下Sternberg（1992）关于如何获得心理学杂志录用的评论是很有帮助的。你的第一个申请不应旷日持久，而应具有很高的获得资助的可能。

以第一个资助项目为基础，在相关的杂志上发表2~3篇研究文章。这样，在寻找第一个主要的外部基金时，你就会处于一个有利地位。我的建议是，尽快写出第一个资助申请（如花费一二周的时间）。首先，你的研究应建立在已有的学术基础上，这个基础既是理论性的也是实验性的。考虑到你有了现存的文字材料，那奖金申请书的前半部分（目的、文献综述、背景研究和假设）会完成得很快。更多的时间可花在研究申请部分——虽然我猜想大多数研究已在你的“小本子”上勾画好了。你需要快速完成第一项基金申请，原因是你很可能在第一次尝试时得不到资助。

坚持的重要性

大约80%的优质基金申请起初都会被联邦机构和基金会拒绝，因此遭到拒绝是基金申请中不得不面对和忍受的事实。更为不幸的事实是：只有20%的基金申请在初次递上去的时候就被接纳。那就是说，大部分基金申请都是呈递过多次的。例如，在NIH，大约75%的基金申请是在第二轮或第三轮才被接受的。这意味着，虽然在某种程度上你失败了，但你会得到有价值的反馈，这对你研究计划的修改和完善是非常必要的。如果你诚恳全面地接受这些批评，你第二次成功的可能性将大大增加。如果不是这样，第三次也许就是你的“最具魅力的时刻”。因此，在成功的基金获得者当中，往往心照不宣的最大秘密就是：失败乃成功之母。坚持！

定位多个基金来源

虽然文章一次只能发表在一本杂志上，但基金申请可以并且也应该同时发往多个基金申请机构。例如，同样的基金申请（只需很少修改来迎合特定的格式要求）可以同时 NSF、NIH 和一些私人基金上评阅。当然，如果从一个机构得到了肯定的评价，那么基金申请应从其他机构撤回。有时候，你的基金申请可以一部分由一个机构资助，而另一部分由另外机构资助。当然，这要具备取得意外之财的运气了。合适的时间和合适的地点是赢得基金的重要因素。

写作风格的重要性

一项研究申请应使读者不需要冥思苦想就能理解和评价你的思想。为了帮助评阅者理解申请，你应引导他们从最初的一般假设起，逐渐过渡到最后关于研究的重要性。如果你确实想让评阅者阅读、评价你的研究计划时不感到过分与无聊，你必须激发评阅人的兴趣，触动评阅人的敏感神经。最初，你的标题和摘要可能是吸引读者注意力的最重要因素。你的评阅人可能做着相似的研究，或发现你的标题和蕴涵的思想很吸引人，虽然他可能一点都不知道背景文献或采用的方法。

把申请当作针对一个不太了解你论题的人，而不是熟知你领域的人来写，这样你就可以使你的文章更浅显易懂。你必须引导评阅人的思想，从一般陈述到相关的论题和文献，再到特定的研究假设、设计和步骤，最后到数据分析和数据的显著性。准确的话语、有趣的风格、恰到好处的细节描述，最有可能增强文章的可读性。一个可读性高的研究计划会吸引观众的眼球，而一个乏味模糊的提案则会招来负面评价。美国科学院院士、美国发表论文最多的科学家之一、化学家Ernest Eliel曾经说过：“并非我比其他人更适合当科学家，而恰恰因为我是更好的作家。”你的基金申请要想吸引和影响目标观众——评阅人委员会，就必须有很高的可读性。

沟通的技巧，包括口头上和笔头上，通常不是简单地在你的科学训练领域练就的，而是你接受教育的早期就练成了——在小学、中学、大学的作文课上，或者给杂志投稿的过程中。我发现很多的心理学家在接受研究生教育期间和以前，没有学会提高自己的写作技巧。这些技巧对申请研究基金来说非常重要，毕竟文字准确、信息全面、有趣可读的文章容易获得成功。

推 荐 阅 读

本章列出了一些关于基金申请的窍门，更为详尽的描述可以在最近出版的精彩文章中找到，如Ries和Leukefeld的《申请研究基金》（*Applying for Research Funding*, 1995）以及Locke、Spirduso、Silverman的《研究计划》（*Proposals That work*, 1993）。这些书主要描述如下的问题：怎样开始申请，写什么和什么时候写，检查疏漏，写特定的内容，做决定的过程。Ries and Leukefeld（1995）书中很多部分都会对你的基金申请准备和操作有所帮助。

最后，全部的过程可浓缩成三个简单的原则：（1）形成一个你引以为自豪的思想；（2）集中精力写出你的思想和研究启示；（3）坚持，直到其他人看到它的优点。那样的话，你就会获得巨大的满足：赢得你的第一份研究基金。

参 考 文 献

- American Psychological Association. (1994). *Publication manual of the American Psychological Association*. (4th ed.). Washington, DC: Author.
- Locke, L. F., Spirduso, W. W., & Silverman, S. J. (1993). *Proposals that work*. Thousand Oaks, CA: Sage.
- Ries, J. B., & Leukefeld, C. G. (1995). *Applying for research funding*. Thousand Oaks, CA: Sage.
- Sternberg, R. J. (1992). How to win acceptances by psychology journals: 21 tips for better writing. *American Psychological Society Observer*, 5(5), 12.



第 28 章

跨文化研究方法

KWOK LEUNG
FONS VAN DE VIJVER

世界正逐渐成为一个真正的地球村。商业的全球化导致不同民族的人工作在同一个屋檐下。移民模式已迅速改变了相对单一的社会民族结构。例如，人们预测，在不久的将来，西班牙裔将成为美国最大的少数民族。此外，诸如全球温室效应、武器限制等挑战也要求许多国家开展合作。

全球化趋势导致不同文化间的接触急剧增多。因此，目前我们面临的迫切任务就是理解不同文化的异同、把文化偏见与真正的跨文化差异区分开来。事实上，心理学研究就可使我们看到这些差异。本章的目的就是针对跨文化研究方法中的有关问题，进行全面的综述。

跨文化研究的特点

在“真正的实验”中，参与者被随机分配到不同的实验条件中，如实验组和控制组（参看本书Wampold所著的第5章）。但是，文化是跨文化研究的一个自变量，我们无法将被试随机分配到不同的文化中去，比方说，不能将一个日本被试分到美国组中。文化是个不为研究者所控制的变量，此特性使研究者非常难以估计跨文化研究中的因果关系。当我们发现两种文化组有所不同时，很难对导致文化差异的真正原因下结论。例如，假如我们从推理测验中发现，美国大学生比越南农民的成绩高。对这种成绩的差异，我们可以有无数种解释：可能因为越南农民所受的教育比美国大学生少，或者他们对这类推理测验不如美国大学生熟悉，这些都导致他们成绩较差。假如是测验的熟悉性引起这些差异的话，使用越南农民熟悉的题目就是一个解决方法，这会使他们的成绩有所提高。对跨文化研究者而言，为跨文化差异寻找一个较为可靠的解释就是一个主要的挑战。

跨文化研究者的另一个挑战是“文化”这个概念过于空泛，很难定义。如果我们观

察到，与美国人相比，中国人更为强调学习成绩中努力的重要性（Hess, Chang, & McDevitt, 1987），那我们可能将其归因为文化差异。可是，文化这个概念包含太多的因素，以至于用“文化”来解释这种差异几乎毫无意义，何况这种差异可能也有别的原因，尤其是社会化、父母教养方式、教育体制的影响。因此，我们需要将文化这个概念分解成一组因素，并且确定哪个因素与我们所了解的文化间差异有关（Whiting, 1976）。

文化对等

绝大多数跨文化研究都至少包括两个文化组，但也有一些研究采用单一文化组，将之与以往研究结果进行跨文化比较。当我们试图比较两种以上文化时，可能会采取两种形式：结构导向型或水平导向型。结构导向型研究调查不同变量间的关系，试图鉴别不同文化中变量关系的异同。例如，在不同文化中，价值观的结构有什么相似点？与此相反，水平导向型研究关注变量在不同文化间的量的差异。例如，文化A中个体的自尊水平是否高于文化B中的个体？

在这两种研究中，对等是进行跨文化研究的基础，因为我们不能拿苹果和橘子进行比较。意义相似称作概念或结构对等，这是结构导向型研究的关键因素。同一概念在不同文化中可能有不同的意义。比如，抑郁对中国人和美国人而言可能有不同的含义。对美国人而言，抑郁仅指心理上的所遇到的问题。而中国人则喜欢用生理症状来描述心理问题，通常称为躯体化过程。对中国人来说，抑郁还包括肉体上的痛苦。

许多方法可保证概念对等。我们可以考虑采用测量概念的工具来测量心理特性。例如，我们用同一量表来测量中国人和美国人对“抑郁”一词的理解。当结果在心理测量特性上表现一致时，我们可以认为“抑郁”这一概念在两种文化中是对等的。实际中的例子是，我们用Cronbach's alpha系数测量其内在的一致性，或用因素分析发现相似的因素结构。另一种可确定概念对等的方法，即去发掘不同文化中抑郁与其他变量类似的相关方面。例如，在两种文化中，如果抑郁都与不良生活事件相关，我们就可以更为肯定地认为，抑郁在两种文化中具有相似的含义。一般来说，如果抑郁的测量显示，我们期望的结果与一系列测量具有相关性，那就支持了概念对等。在研究概念对等时，我们可以使用各种统计技术，如回归分析、多维量表法、因素分析、结构协方差分析/结构方程。

在水平导向型研究中，我们希望在概念对等的基础上，找出不同的文化组在某变量上的平均分是否存在差异。但不幸的是，即便概念是对等的，我们仍不能确信，在不同文化之间，由同一量表得到的数据可以直接进行对比。例如，当我们测量身高时，一个组使用英寸作为单位，而另一组使用厘米作为单位，虽然两组的确是在测量同一个概

念,但数据不能直接进行比较。同理,假设我们测量发现,与美国人相比,中国人和日本人报告出较低的自尊水平,如果没有进一步的证据,我们就不能确定该差异的真实性。因为,选项设置可能使中国人和日本人的回答产生偏差,因为他们更倾向选择等级量表的一部分,或者他们可能故意报告较低水平的自尊,因为谦虚的社会规范不鼓励人们吹嘘自己的能力。

在比较从不同文化中测得的数据前,我们需要确立数量等价或者数据可比较性,身高和体重就是数量等价的变量例子。对心理结构来说,确立数量等价并非易事,我们进行跨文化比较时必须注意这个问题。当我们宣称在某个变量上,一个文化组高于另一个文化组时,必须表明我们已有较满意的数量等价水平。读者如果想更多地了解建立数量等价的一般过程,可阅读本章最后一节推荐的读物。

跨文化研究的设计

我们通常很容易发现不同文化之间存在的显著差异,但却很难解释这些差异。不同文化组在多方面都有所不同,这些不同又可能是观察到的差异的原因。在进行跨文化研究设计中,设计者必须努力提高文化差异的可解释性。一个解决方法就是增加协变量,像教育水平这类未受关注的变量有可能会产生文化差异。当找到跨文化的差异时,设计中的协变量有助于评估我们提出的解释,排除其他的解释。Earley (1989) 关于“社会性懈怠”的研究就是一个很好的例子。“社会性懈怠”是指人们在团队中工作时付出的劳动要比他们在单独做这项工作时付出的少。研究发现,与中国被试相比,美国被试显示出更高的社会性懈怠水平。于是,个人主义—集体主义的测量就成了本研究的协变量。个人主义是指一种不怎么强调个人小圈子(如家庭)的倾向,而集体主义与其相反,很强调团体利益(Hofstede, 1980)。在对个人主义—集体主义方面的跨文化差异进行控制后,社会性懈怠的文化差异就消失了。协变量分析为研究者使用个人主义—集体主义来解释社会性懈怠提供了强有力的证据。

Leung & Zhang (1996) 指出,许多研究的出发点都是西方和非西方文明的比较,但有些调查研究并不符合当地实际,这可能是研究者文化背景的影响。对同一项研究,不同文化背景的研究者可能使用不同的设计,因此得到不同的结果。所以,研究者要避免文化因素对研究问题的影响。为此,可采用两种方法:

第一是去中心化法,即在一项研究的设计和概念中吸纳不同的文化观点。例如, Schwartz (1992) 在研究价值观的结构时,鼓励来自不同文化的研究者将其特有文化价值观加入到泛文化价值观的集合中。

第二种是会聚法。这种方法的基本思路是设计一个研究,与现有研究之间的文化差距尽可能多一些,以此来看研究结果能否与现有研究的结果会聚。如果可以会聚,则说

明现有研究的文化背景没有影响结论；如果发现不同结果，则说明现有研究的文化背景使结果产生误差。有一个很好的例子能说明这种方法：Bond和他的同事设计了一个完全基于中国人价值观念的价值观调查，并在22个国家实施了（中国文化联盟，1987）。他们研究发现，有三个因素与Hofstede（1980）所确定的因素相一致（Hofstede使用的是完全基于西方文化而设计的测试工具）；同时还发现一个新因素——儒家工作动力学，该因素与国家的经济发展高度相关。

对文化的抽样

对验证跨文化研究假设来说，选择文化相当重要。研究者经常采用系统抽样，即以某种理论为依据选取文化。Berry（1966）的经典实验是一个有关系统抽样的例子。研究对象分别是以农业为生和以狩猎为生的两组被试。研究者假设：农业社会对个人施加更强的顺从压力，因此，该社会的个体更多地表现出场依赖性特点（场依赖性是指知觉物体时，会更多地受物体所在背景的影响）；而狩猎社会鼓励成员自主，因此表现出较低水平的场依赖性。对这两个文化组进行系统抽取，以验证这个假设。

随机抽样，即从许多文化样本中进行随机抽取，常用于估计一个心理构件的普遍结构或一个泛文化理论。但真正的随机样本几乎不可能得到。不过，几个大规模研究的抽样也可近似地看作随机抽样，如Schwartz（1992）抽取了20种文化样本来研究人类价值观的结构。

被试的抽取

要想进行有效的比较，不同文化组的被试在背景特征上应该是相似的。不然的话，研究者就会很难确定，观察到的差异究竟是因为文化上的差异，还是因为样本误差。假如我们把一种文化中的文盲和另一种文化中受过高等教育的人进行比较，教育水平差异就能更好地解释结果，而非他们文化的差异。根据人口统计学特征，匹配样本是解决这个问题的一种方法，这样能消除样本的差异（详见本书McCready所著章节）。研究者经常假设，来自于不同文化的大学生具有相同的人口统计学特征，因此常选取他们作为被试作比较。利用同样的思路，Hofstede（1980）为了减少环境变量的影响，从一个分布在40多个国家的跨国公司的分组织中抽取被试。

有时候，因为实际操作的问题或被试在人口统计学特征上明显的跨文化差异，所以几乎不可能匹配不同文化的样本。解决的方法是将相关的人口统计学变量作为协变量。例如，在一个比较美国、澳大利亚和中国香港对行为不良青少年的研究中，研究者发现，中国香港被试的父亲们的教育水平明显低于美国和澳大利亚（Feldman, Rosenthal,

Mont-Reynaud, Leung, & Lau, 1991)。为克服这个问题,研究者以三种文化团体中的父亲教育水平为协变量,进行协方差分析。

数 据 收 集

在我们用一种工具,从不同文化团体中收集数据之前,需确定它确实适用于每个团体(详见本书Ponterotto所著章节)。比如,测量的某些项目可能不适用于某些文化。当在中国使用明尼苏达多项人格量表(MMPI)时,研究者发现,有些题目用中文表达根本没意义,因此进行了修订(Cheung, 1989)。假如太多的题目不合适,那么就需要研制一个全新的量表。Church(1987)发现,西方人格测量工具测不出菲律宾人的人格结构,因此,他建议应采用新的人格测量工具,进行测量。

在进行跨文化研究时,我们有时也需要将测量工具翻译成为另一种语言。翻译一回译法是一种广泛使用的方法(Brislin, 1980)。该方法是将某种语言的量表翻译成另一种语言,然后,再由另一个翻译者把它翻译回原来的语言,从中进行比较。这种方法通常是可行的。在研究者不太精通研究对象的语言的情况下,回译可帮助研究者掌控翻译的准确性。但这种方法也有缺点,它可能会变为修饰过度的语言,即翻译者翻译得非常好,但难以理解。在测试条目中包含难译的成语时,这个问题尤其严重。

Werner & Campbell(1970)提到一种叫“去中心化过程”的翻译方法,即对原文和译文同时进行校正。去中心化过程的目的是为了在一种语言中一成不变地再现原文的工具,而是为了同时增强所有不同语言译文的可读性和会聚性。

在我们将一种测量工具用于不同文化时,要注意实验者和访谈者的人格特征。主试或访谈者可能会影响到被试的反应,特别是当主试与被试的文化背景不同时。曾有学者系统地研究过智力测验中因主试与被试之间种族不同而产生的误差。另外,主试与被试之间的交流也是一种可能的误差源。主试可能会在无意中给出一些信号或暗示,如外表或动作。最后,刺激和反应过程也可能是误差源,如被试对问卷不熟悉,把它当成是具有对错之分的考卷。Deregowski和Serpell在1971年的一项研究说明了采用熟悉刺激物的重要性。在一个实验条件中,他们要求苏格兰孩子和赞比亚孩子将缩小的动物和摩托车模型进行排序;在另一个实验条件中,他们要求孩子们对那些模型的照片进行排序。结果发现,在对模型进行排序时,没有发现文化差异,而在对照片进行排序中,苏格兰孩子的成绩明显高于赞比亚孩子的成绩。

跨文化研究的四种常见类型

根据研究方向是试探性的还是假设验证性的、是否需要考虑背景或文化因素，我们可以将跨文化研究分为四大类。前两类重点放在假设检验上。普遍性研究力图验证在一个文化群体（通常是西方）所取得的研究结果能否推广到其他（西方或非西方）文化群体。一般而言，这类研究很少涉及地方文化因素。例如，Schwartz（1992）从许多国家采集数据，用来研究全球人类的普遍价值观结构。

在理论导向研究中，文化因素被看成是理论框架的核心。研究者关注文化变量与心理学理论的关系，并采用假设检验的方法研究这些关系。研究过程包括文化抽样和证实存在文化差异。研究者根据不同的研究焦点抽取不同的文化——即可提前预期差异存在之处。这些研究焦点均根据现有理论框架提出。例如，Leung（1987）预测，在冲突情境中，集体主义文化影响调解策略的使用。研究者选取中国人和美国人两个样本，代表集体主义的两种不同水平。正如预测，研究证实了在冲突情境下对调解策略的偏好存在文化差异。上面提到的Berry在1967年的一项关于场依赖性的研究是另一例子。

在接下来介绍的两种跨文化研究中，假设检验法并不重要。在心理差异研究中，研究者会测量至少两种文化，重点检验不同文化团体测量数据的平均数、标准差、信度系数及其他心理测量学指标。在这类研究中，通常没有什么大理论可预测跨文化的差异。因此，研究者在观察到文化差异后再去寻找解释。例如，Guida and Ludlow（1989）比较了美国孩子和智利孩子焦虑问题的测验。研究者发现，在大班和中班的被试中，美国孩子测验焦虑的程度比智利孩子低。就像通常这类研究所做的那样，实验者对观察到的差异进行事后解释。

最后一种跨文化研究是外部证实。该类研究基于某个与文化相联的因素，探究跨文化差异的意义和差异产生的原因。它并不需要一个事先的、特殊的假设；取而代之的是，试探性地采用大量与文化相联系的变量，来解释跨文化差异。这种研究不是为探究概念或者数量上的对等，而是在经验基础上寻求对跨文化差异的解释。例如，Bond（1991）研究了许多民族的价值观和多种健康测量之间的关系，发现价值观与一些和健康有关的变量之间存在关联。

总 结

在跨文化研究中，至关重要的是解释研究发现的意义，但这种解释往往模棱两可。这是因为从本质上说，跨文化研究和真正的实验研究有所不同。对于跨文化的差异，有很多具有代表性的解释。在我们看来，最好的方法是事先阐明许多相互对立的假设，然

后在实验设计中排除其中的一些解释。与常规的单文化研究相比,跨文化研究需要采用更高水平的混合方法。虽然这种研究很复杂,但在这个日益全球化的世界中,我们无法忽视文化上的异同。如有仔细的计划 and 合理的操作,跨文化研究不但极具价值,而且还可带来不同文化间丰富多彩的交流。

推 荐 阅 读

如果想详细了解跨文化研究的各类方法,请参看Brislin、Lonner and Thorndike, (1973); 如果想详细了解跨文化研究的方法与数据分析,请参看Van de Vijver and Leung (in press); 如果想了解现场调查研究方法,请参看Lonner and Berry (1986)编辑的书,一定很有作用。

参 考 文 献

- Berry, J. W. (1966). Temne and Eskimo perceptual skills. *International Journal of Psychology*, 1, 207-229.
- Bond, M. H. (1991). Chinese values and health: A cross-cultural examination. *Psychology and Health*, 5, 137-152.
- Brislin, R. W. (1980). Translation and content analysis of oral and written material. In H. C. Triandis & J. W. Berry (Eds.), *Handbook of cross-cultural psychology* (Vol. 1, pp. 389-444). Boston: Allyn & Bacon.
- Brislin, R. W., Lonner, W. J., & Thorndike, R. (1973). *Cross-cultural research methods*. New York: John Wiley.
- Cheung, F. M. (1989). A review on the clinical applications of the Chinese MMPI. *Psychological Assessment*, 3, 230-237.
- Chinese Culture Connection. (1987). Chinese values and the search for culture-free dimensions of culture. *Journal of Cross-Cultural Psychology*, 18, 143-164.
- Church, T. A. (1987). Personality research in a non-Western setting: The Philippines. *Psychological Bulletin*, 102, 272-292.
- Deregowski, J. B., & Serpell, R. (1971). Performance on a sorting task: A cross-cultural experiment. *International Journal of Psychology*, 6, 273-281.
- Earley, C. (1989). Social loafing and collectivism: A comparison of the United States and the People's Republic of China. *Administrative Science Quarterly*, 34, 565-581.
- Feldman, S. S., Rosenthal, D. A., Mont-Reynaud, R., Leung, K., & Lau, S. (1991). Ain't misbehavin': Adolescent values and family environments as correlates of misconduct in Australia, Hong Kong, and the United States. *Journal of Research on Adolescence*, 1, 109-134.
- Guida, F. V., & Ludlow, L. H. (1989). A cross-cultural study of test anxiety. *Journal of Cross-Cultural Psychology*, 20, 178-190.
- Hess, R. D., Chang, C. M., & McDevitt, T. M. (1987). Cultural variations in family beliefs about children's performance in mathematics: Comparisons among People's Republic of China, Chinese-American, and Caucasian-American families. *Journal of Educational Psychology*, 79, 179-188.
- Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Beverly Hills, CA: Sage.
- Leung, K. (1987). Some determinants of reactions to procedural models for conflict resolution. *Journal of Personality and Social Psychology*, 53, 898-908.
- Leung, K., & Zhang, J. X. (1996). Systemic considerations: Factors facilitating and impeding the development of psychology in developing countries. *International Journal of Psychology*, 30, 693-706.
- Lonner, W. J., & Berry, J. W. (Eds.). (1986). *Field methods in cross-cultural psychology*. Beverly Hills, CA: Sage.
- Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 25, pp. 1-65). Orlando, FL: Academic Press.
- Van de Vijver, F. J. R., & Leung, K. (in press). *Methods and data analysis for cross-cultural research*. Beverly Hills, CA: Sage.
- Werner, O., & Campbell, D. T. (1970). Translating, working through interpreters, and the problem of decentering.

In R. Naroll & R. Cohen (Eds.), *A handbook of cultural anthropology* (pp. 398-419). New York: American Museum of Natural History.

Whiting, B. B. (1976). The problem of the packaged variable. In K. Riegel & J. Meacham (Eds.), *The developing individual in a changing world* (Vol. 1, pp. 303-309). The Hague: Mouton.



第 29 章

在研究中应用理论

——科学理论和科学研究的相互影响

Charles J. Gelso

提问一个典型的心理学专业本科生（通常是好学上进的学生），心理科学由什么组成？几乎所有人的回答中都有实验，而理论却鲜有人提及。事实上，我必须承认，当我还是研究生时，我认为科学研究与实验大致就是一回事。现在唯一能让我安慰的是，我并不是唯一持这种错误观点的人。20世纪60年代激进行为主义学派在美国占统治地位，致使当时许多学生甚至经验丰富的学者都将实验与科学研究等同起来，并认为理论是不必要的，甚至会阻碍科研工作。当时，大家的研究都遵循行为主义原则。

到了20世纪90年代，心理系已发生了翻天覆地的变化。即使这样，仍有残留的少数观点认为，理论在科学进程中是二等公民。然而，在这一章里，我要说的是，理论和实验之间存在着深刻而必然的联系，轻视任何一方都会使科研成果濒于枯竭。因此，二者都是科学的必要元素。如果没有可控的经验性研究，那么科学仅仅是未经检验的想法和偏见的集合而已，难以想像这些是科学的结果，例如，魔法和占星术就不是科学。如果没有理论，科学仅仅是一堆毫无关联的现象（有人称之为事实）组合，而不是对心智有意义的理解。

人们不仅常常弱化理论在心理科学中的作用，而且也常常误解理论的组成。因此，在本章的开始，我将从以下这些方面来总结：如何看待理论的定义、理论的组成元素、怎样的理论才算是好的科学理论。我相信，先弄清楚这些，将对理解理论有所帮助。然后，我会举例说明，在科研中如何应用理论和实验，以及它们是如何互相渗透、互相促进和相互关联的。

什么是科学理论

理解理论与实验之间联系的最大障碍之一就是曲解理论本身的构成。许多学生将理论等同于已提出多年的、庞大的理论体系，如心理分析、行为主义和人道主义。同时，

在实践导向的领域里，人们常认为理论是治疗方法系统，如个人中心治疗、认知治疗、心理分析治疗。我在别处也曾提到，这些有关治疗的大规模理论以及与此相关的人格理论，常常“和生活、人类哲学及信仰问题联系在一起（这些都是不可检验的）；因此，这些理论或多或少是在不关注心智的可测性的前提下发展起来的”（Gelse, 1991, p.212）。这些不寻常的理论总是难以被驳倒。即使它们在某些方面可以被证明不成立，人们仍然不能证明整个理论是错误的。总之，这些理论系统对科研工作并无太大用处，因为它们不能产生可以验证其有效性的新研究。

更为有用的是被称作微型理论的理论，它们可以是更广理论系统的一部分，也可以是从现存理论系统剥离出来的理论陈述。一个扎根于更广泛的理论的例子，是卡尔·罗杰斯有关有效的心理咨询和治疗的充分必要条件的著名论断。罗杰斯经过多年的实验和实践，提出了一套在当时看来大胆甚至是卤莽、但后来被证明极具启发性的理论。他指出，要对患者做出建设性的改变，需要满足一些针对患者和治疗师的充分必要条件，其中三种条件（治疗师以移情的态度理解患者、对患者无条件的积极关注、和患者保持一致）在过去30年里引发了大量的研究和理论。这些理论后来被纳入罗杰斯的患者中心疗法理论之中，接着又被纳入了更大范围的理论——人类行为和心理咨询的人本主义的研究。

上文曾提过，这些微型理论不一定要纳入更大的理论系统之中，它们能以独立的姿态存在。我在研究治疗者反移情管理时提出的理论就是这样的例子（论述见后）。目前在治疗领域，Bordin（1979）有关治疗中劳动联盟组成和角色的微型理论引出了大量的研究，表现出一个好的理论所应具有启发价值（好理论的标准见后）。我们必须注意到，即使是以独立姿态出现的小理论也与大理论体系有着很大程度的关联。我的反移情管理理论和Bordin（1979）的劳动联盟理论都在很大程度上属于心理动力学的范畴。

那么什么是理论呢？最自由化的定义是：

一个理论可被看作是一系列通过猜测、假定或事实证明来揭示它们之间相互关系的两个或多个结构（抽象）。理论命题定义了结构（现在称为“变量”）之间的关系，当这一理论命题不再有争议，并被认为属于相关领域的知识时，它便成为事实。理论在抽象、客观-主观、现实主义-理想主义、观察性、正式-非正式等方面具有不同的水平上变化（Rychlak, 1968, p.42）。

说到底理论是对变量间关系的猜想。从这个角度看，我们所有的实验和工作都含有理论，它是实验变量间关系的表达。但正如上文指出的，理论有正式和非正式之分。所谓非正式理论指其表述不明确，并且不以得到“合乎逻辑的关联知识”（Rychlak, 1968, p.35）为目的。作为科学家，我们的目的是要提出明确的理论，让经验性的研究对其进行检验，并从而刺激进一步的研究。

理论的功能和性质

我认为理论要想更具科学价值，就必须超越命题的水平。一个理论必须告诉我们为什么变量之间被预期会相互关联和影响。也就是说，在预期变量之间如何相互作用的背后必须要有站得住脚的理由。

将任何一个理论看作或多或少地服务于某些功能，这也是有用的。Rychlak 讲述了理论的四种功能：描述功能、界定功能、启发功能、整合功能。这四种功能是值得仔细考察的。

首先来看描述功能，所有理论都需要描述一些现象，将现象的发生条件描述得越详尽，就越接近科学意义上的解释。详尽的描述可以看作是等价于对原因的解释。好的理论致力于解释现象发生的原因，它们具有很强的解释力（explanatory power）。

理论也具备界定功能。这意味着通过理论的表述，人们就能知道条件和结论的边界，从而能了解理论适用的范围。这样的界限很有必要，它们对要考察什么起着导向性作用。

也许最重要的是，理论能够引发新的想法和需要进一步的检验。具有高度启发性的理论能刺激新的研究，促使对理论进行进一步的验证。不具备启发性的理论将会停滞不前，而且从本质上是讲是没有科学价值的。可以看出，这一点奠定了理论与经验研究之间的整合关系。

最后，理论具有整合功能。理论寻求以一种连贯的、统一的方式来聚合命题和变量。换句话说，好的理论把分散的、当时看来相互矛盾的事实统一到一个具有贯通性和高度内部一致性的画面中来。

有一种整合功能尤其值得人们注意，因为它很重要，也因为它在当前处于被遗忘的角落，那就是精简原则（Parsimony）——一种理论综合。这一原则可能被认为是理论家们在解释所考虑范围之内的现象时尽可能少地引进理论结构和命题。在最近一次对精简原则的讨论中，我问在座的8位心理咨询专业的毕业生，是否曾经听说过奥肯剃刀（Ockham's razor）。结果发现，只有一名同学的答案是肯定的。这一词条的起源可以追溯到14世纪英国哲学家、来自奥肯村的威廉。这意味着，在任何一个理论中，如果存在不为解释所考虑范围内现象所必需的理论结构和命题，那么这些结构和命题，就会成为累赘。这些部分需用剃刀剃除。现代的心理学家在创建自己的理论时也要“剃刀在手”。

好的科学理论

通过前面对理论功能的讨论，我们已经看到了一个好的理论需要具备的一些要素。现在我们就来总结一下并增加一些其他因素。

正如前面所述，（1）好的理论必须有内在的一致性，且综合、精简；（2）必须有

高度的解释力；（3）在一个界定清晰的领域内，好的理论应该是全面的，也就是说，它说明了该领域内所有的重要关系；（4）一个理论要具有最佳的科学价值，它必须清晰地、明确地解释某一领域的变量，这样，它的命题才具有高度的可检验性，并且可以被用来证伪。如果一个理论不具备被证伪的可能，它的科学价值也就很有限了。事实上，每个假设驱动的研究都应将理论置于被证伪的可能下。在这个意义上，面临被推翻的理论就是一个好的理论！回顾一下前面提到的关于人格和治疗的理论，它们的局限性之一就是没有被证伪的危险。就像Mahrer（1988）指出的，这使得检验这些理论的工作几乎是徒劳的。

最后，正如我们前面所讨论过的，一个好的理论其底线是它具有启发价值，即一个好的科学理论应能引起人们的质疑，进而引发经验性研究和新理论。通过前面对理论好坏的讨论，实验之于理论的重要性已经凸显出来。为阐明理论和实验二者的关系，我将在下文引入“科学思想是如何产生的”这一问题，并阐述何谓“科学循环”，即理论和实验之间互相促进的关系。最后，我将论述并举例说明在实验过程（产生想法、假设、解释现象的探索）中如何使用理论。

下文中的例子大部分来源于我对反移情作用的研究和理论结构。所谓反移情作用，即指咨询人员或精神治疗医师对患者或病人的情感反应，这种反应以治疗师的经历为基础，常与其经历的早期、未解决的冲突有关，比如与父母之间的冲突。

研究想法的产生

科研工作中的问题是如何产生的？首先，让我们通过下面的例子来看一下实验问题的来源。我们在马里兰大学开展反移情作用研究的初期，有一位名叫安·皮博迪的研究生对心理咨询师的移情能力与反移情（叫对患者的情感抽离）倾向的关系感兴趣。我们检查了关于这两个变量的精神分析理论，并在此理论以及作为咨询师的体验基础之上，我们假设，面对表现出攻击行为和诱惑行为的女性患者所陈述的内容男性心理咨询师的共情能力越强就越不会回避（从理论上讲，这些行为对心理咨询师具有胁迫性）。我们认为，共情能力强的心理咨询师同样也会有反移情体验，但问题的关键在于，他们更可能意识到这些体验，但不会对病人施展出来。我们给心理咨询师播放由女演员扮演的患者的录音，在录音暂停的时段，观察心理咨询师的反应，实验结果支持了我们的假设（参照Peabody & Gelso, 1982）。当然，在这里我们关注的是，上述研究的假设来自于理论（以及我们的临床经验）。虽然没有理论精确地表述了我们的假设，但这些假设可以由理论逻辑推论出来。如果没有移情和反移情的理论，就不可能有这些假设。

我们也会问：理论上的想法源自哪里？在最近的研究工作中，我和我的同事认为，虽然所有的心理咨询师都体验过反移情，但治疗中最重要的事情是如何有效地控制或管

理这种感情。如果管理得当，它将会对治疗有益，反之将会阻碍治疗。基于这种理论，我们提出了心理咨询师有效管理反移情作用能力的几个成分。这个理论吸收了Penbody和Gelso的研究成果，包括在此基础之上的研究和其他无关的研究。它假定，反移情管理能力包含5个相关因素：心理咨询师的自我洞察力（Peabody & Gelso, 1982; Robbins & Jolkovski, 1987）、心理咨询师的共情能力（Peabody & Gelso, 1982）、心理咨询师对焦虑的管理能力（Hayes & Gelso, 1991）、心理咨询师的自我整合能力（McClure & Hodge, 1987）、心理咨询师的概念化能力（Robbins & Jolkovski, 1987）。

我们建构理论时，就像我们考察研究理论从哪里来一样，不单依靠研究，也靠个人的或临床经验。这种个性化不可避免，它构成了理论—研究—理论—研究的基本链接。

科学工作的循环

上面我考察了理论是如何引发研究，研究又是如何引发并完善理论的；而且我们也已讨论了理论的定义以及好的理论的标准。现在，我们可以来探讨我所认为的科学核心：理论和研究如何互惠式的相互关联。它们之间是很深入的协同关系。它们并肩合作，为的是生产出最好的东西——好的科学。Stanley Strong（1991）在讨论“科学工作的循环”（p.208）时，曾细致地描述过这种协同关系。和许多前辈一样（如Meehl, 1954; Reichenbach, 1938），他考虑了科学产生的两个环节：发现和检测（也有人称其为论证），认为科学工作由这两者的循环构成，他是这样说的：

在发现环节，科学家根据观察到的现象提出和建构动力学概念。在这个环节中，科学家尽力运用所有的想法、观察、直觉以及创造力。在概念提出后，科学家就想方设法将概念与可观察的现象联系起来，并据此提出如何测量动力学的特征。有了理论以后，科学家就进入检测环节，他们进行一些新的观察去检测理论的假定。不可避免地，有些观测到的现象不能由提出的理论来解释。基于更多的观察结果、思想和直觉，科学家回到发现的环节，去修改理论或提出新的理论。

当然，也许有人会补充说，这个循环是无穷无尽的；理论会被不断地修正，最终被其他理论所代替；新理论继续指导研究，不断被检验、修正，如此等等（我必须指出的是，Stanley Strong对科学的观察具有特定的哲学立场，他称之为伽利略模式。这也是为什么他强调对深层的动力学因素加以理论化。Strong的观点可广泛地用于科学研究）。

一些研究者也许会认为，理论与研究之间并不存在这样的联系和循环。比如说，经过详尽的分析，Mahrer（1998）认为，假设-检测性研究（用以产生和检测理论）在心理学研究领域内的成果并不多。他认为，虽然假设-检测研究在心理学研究中占支配地位，但它却阻碍了人们对新理论的发现，而研究者真正需要的是发现—修正模式。这样，研究者就可摆脱假设所带来的偏见和制约，Mahrer相信这会导致新认识的出现。

科学家当然有自由进行不以检测假设为目的的研究，因为假设检验并不是科学的全部。另外，在当前对心理治疗的研究中，发现-修正模式非常有效。我必须声明，尽管他们的理论常常是非正式的，但研究者在遵循这一模式作研究时也不能脱离理论，因为这些理论会在研究的每一步骤中指导他们的工作。在研究过程的某些点上，发现-修正模式可能是最好的，比如说，当理论和假设检验似乎要将我们带入死胡同、此领域似乎没有什么新鲜见解时，发现-修正式的研究可能正是我们需要的灵丹妙药。即使是在这种情况下，随着以发现为导向的研究所带来的成果不断累积，人们必然将其用于创造、完善新的理论，而这些理论也会引发一些接下来可以检验的假设。如此一来，以发现为导向的研究便可看成是发现环节的一部分：发现的关系有助于形成理论，这些理论由检验或论证而得到考察。

对“在研究中运用理论”的进一步讨论

在这一章中，我主要集中探讨理论和研究在科研中的相互作用。心理学研究历来有忽略理论作用的传统，所以我想对在研究中如何应用理论这一点作额外的阐述。总的来说，理论在实验的每一个过程中都是有用的，但是在以下三个过程中尤为重要，起到了核心作用：（a）思想的产生；（b）假设的提出；（c）对结果的解释。任何经验性研究在其初期，都要产生研究的思想和形成假设（至少是研究的问题）。最近两个关于反移情作用的实验（Gelso, Fassinger, Gomez, & Latts, 1995; Hayes & Gelso, 1993）可以作为例子。在这组实验中，让接受培训的心理咨询师观察一些由演员扮演的、不同性取向的、患者的录像，以考察心理咨询师的反移情反应。有理论认为，人们（包括心理咨询师）倾向于对同性恋关系比异性恋关系抱有更多的负面态度（也可以认为这一观点来自于一个更大的理论，即人们对于与自己不同的人抱有负面态度）。据此我们假设：当患者是男性同性恋者（Hayes & Gelso, 1993）或女性同性恋者（Gelso et al., 1995）时，心理咨询师可能在行为、感情和认知水平上表现出更明显的反移情作用。这个理论也附带假设“同性恋恐惧”（意思是对同性恋男女抱有偏见态度和刻板的负面印象），这一假设的心理结构是影响心理咨询师对同性恋患者作反移情的中介变量。当接受培训的心理咨询师观看这样的影片时，实验结果表明：对不同性取向的患者来说，心理咨询师的反移情反应并没有差别，即心理咨询师的反移情反应与患者的性取向并无显著关系。但我们发现，心理咨询师的同性恋恐惧与反移情反应有显著的关系，心理咨询师的同性恋恐惧值越高，对男女同性恋者的反移情反应也就越明显。

在Gelso等人（1995）的研究中，我们假设当患者是女同性恋时，女性心理咨询师会比男性心理咨询师表现出更强的反移情反应；但当患者是异性恋者时，男女心理咨询师的反应没有差别（这个实验中没有涉及男性患者的情形，但如果有此情形，我们可以同

样假设当患者是男同性恋者时，男性心理咨询师会比女性心理咨询师有更强的反移情反应）。这一假设是基于以下的理论：男性对男同性恋者会比对女同性恋患者抱有更强的负面态度；女性对女同性恋者会比对男同性恋患者抱有更强的负面态度，这部分是因为患者的“同性相吸”会引起同性别心理咨询师下意识的恐惧。对反移情的认知测量支持了这一假设（即咨询师对影片中的患者使用与性有关词语的回忆错误率）。当患者是女性同性恋者时，女性心理咨询师回忆中的错误比男性心理咨询师的要多；而当患者是女性异性恋者时，两者的表现没有差别。

在上面的例子中可以看到，和大多数心理学研究一样，有些假设得到了支持，有的则不然。正是这些“失败”了的假设促使我们重新探究患者性取向、心理咨询师的性别、心理咨询师对同性恋者的排斥程度与心理咨询师的反移情反应之间的关系，修正先前的微型理论，这些修正后的理论会刺激产生下一步研究的假设。

把“假设”与“理论”、“理论命题”区分开来也许是有益的，虽然事实上它们并无绝对的区别（前面引用的Rychlak在1968的论述就没加区别）。一般来说，理论包括理论命题以及来自于命题的假设。假设一般要比命题更为具体，是经验性研究直接检验的对象（如上述反移情研究表现出来的）。

前面已经指出，当实验者试图解释其结果时，理论会变得格外重要。如果结果与来自一个特定理论的假设一致，那么这些发现当然就可由这个理论解释；如果结果出乎预料时，研究者倾向于在已有的理论中找解释或创造自己的理论，这个新的理论往往可嵌入到更广阔的理论系统中去。在反移情的两项研究中，实验者（Latts & Gelso, 1995; Robbins & Jolkovski, 1987）为了解释实验结果，提出了新的理论，即典型的两个因素的交互作用：（1）心理咨询师意识到反移情的程度；（2）在影响反移情行为的工作中对咨询理论的应用。两组实验结果都表明，如果治疗师能很好地意识到自己的反移情情绪，并在咨询过程中常常使用该理论，那么，反移情行为（回避患者的影响）的水平最低。这个结果是实验者预料到的，也符合已有的理论。这说明，治疗师越能准确地意识到自己的反移情情绪，具备越多的理论知识来应对该反应，那么，他就越不会出现反移情行为。

出乎意料的是，研究者们还惊奇地发现，经常使用理论且对反移情情绪的认知度低的心理咨询师产生最为显著的反移情行为。为了解释这些发现，研究者提出了新的理论：当心理咨询师不能意识到自己的感情时，他们使用咨询理论只不过起到心理防御的功能。他们只是对患者和与患者的关系有清醒的认识，但缺乏在感情上的理解。在缺乏感情理解时应用咨询理论，不会有效地阻止咨询师对患者做出反移情行为。你可以看到，虽然研究者提出了一个新的理论，但这个理论从属于一个更广的理论，即心理动力学理论。

正如我在本章里早些时候提出的，我们不能忽视个人或临床经验对理论和实验的指

导作用。在我们选择理论以提出假设和解释发现时,个人经验和实践中对患者的了解都能起到根本性的指导作用。这些经验必须由理智加以引导和管理(与反移情管理类似),这样它们才能有效地指导我们,这也是研究过程中不可缺少的一环。

总 结

在本章中我们看到,心理科学中理论与实验之间存在着必然的、深刻的、相互促进的关系。本章探讨了理论的定义和组成以及“好的理论”的要素。通过介绍科学循环,阐述了理论和实验之间是如何相互作用的。最后以反移情研究为例,讨论了理论在研究工作的初期和末期的作用。

推 荐 阅 读

对理论—实验循环有兴趣的同学,可以看看以下3本书:Strong(1991, p.208-209)讲述了他在研究中使用理论的实例,让读者看到研究过程是怎样要求他修改、甚至放弃他原有的理论;Rotter(1990)清楚地描述了一个理论构造(他用的例子是内部和外部控制)如何在嵌入到更广的理论框架之中时才能在科学上发挥作用;McClelland(1993)以其对成就需要研究为背景,对理论和实验之间的交互作用做了精彩的总结。他发现,在解释某种行为模式时,应该纳入另外一个心理构造——权力需要。

参 考 文 献

- Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research, and Practice*, 16, 252-260.
- Gelso, C. J. (1991). Galileo, Aristotle, and science in counseling psychology: To theorize or not to theorize. *Journal of Counseling Psychology*, 38, 211-213.
- Gelso, C. J., Fassinger, R. E., Gomez, M. J., & Latts, M. G. (1995). Countertransference reactions to lesbian clients: The role of homophobia, counselor gender, and countertransference management. *Journal of Counseling Psychology*, 42, 356-364.
- Hayes, J. A., & Gelso, C. J. (1991). Effects of therapist-trainees' anxiety and empathy on countertransference behavior. *Journal of Clinical Psychology*, 47, 284-290.
- Hayes, J. A., & Gelso, C. J. (1993). Male counselors' discomfort with gay and HIV-infected clients. *Journal of Counseling Psychology*, 40, 86-93.
- Hayes, J. A., Gelso, C. J., VanWagoner, S., & Diemer, R. (1991). Managing countertransference: What the experts think. *Psychological Reports*, 69, 139-148.
- Latts, M. G., & Gelso, C. J. (1995). Countertransference behavior and management with survivors of sexual assault. *Psychotherapy*, 32, 405-415.
- Mahrer, A. R. (1988). Discovery-oriented psychotherapy research: Rationale, aims, and methods. *American Psychologist*, 43, 694-703.
- McClelland, D. C. (1993). Motives and health. In G. G. Brannigan & M. R. Merrens (Eds.), *The undaunted psychologist* (pp. 129-141). New York: McGraw-Hill.
- McClure, B. A., & Hodge, R. W. (1987). Measuring countertransference and attitude in therapeutic relationships. *Psychotherapy*, 24, 325-335.

- Meehl, P. E. (1954). *Clinical versus statistical prediction*. Minneapolis: University of Minnesota Press.
- Peabody, S. A., & Gelso, C. J. (1982). Countertransference and empathy: The complex relationship between two divergent concepts in counseling. *Journal of Counseling Psychology, 29*, 240-245.
- Reichenbach, H. (1938). *Experience and prediction*. Chicago: University of Chicago Press.
- Robbins, S. B., & Jolkovski, M. P. (1987). Managing countertransference: An interactional model using awareness of feeling and theoretical framework. *Journal of Counseling Psychology, 34*, 276-282.
- Rotter, J. B. (1990). Internal versus external control of reinforcement: A case history of a variable. *American Psychologist, 45*, 489-493.
- Rychlak, J. F. (1968). *A philosophy of science for personality theory*. Boston: Houghton Mifflin.
- Strong, S. R. (1991). Theory-driven science and naive empiricism in counseling psychology. *Journal of Counseling Psychology, 38*, 204-210.
- VanWagoner, S., Gelso, C. J., Hayes, J. A., & Diemer, R. (1991). Countertransference and the reputedly excellent therapist. *Psychotherapy, 28*, 411-421.

